Understanding the performance of healthcare services: a data-driven complex systems modeling approach

Li Tao

Hong Kong Baptist University

Follow this and additional works at: https://repository.hkbu.edu.hk/etd_oa

Recommended Citation
Tao, Li, "Understanding the performance of healthcare services: a data-driven complex systems modeling approach" (2014). Open Access Theses and Dissertations. 89.
https://repository.hkbu.edu.hk/etd_oa/89

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at HKBU Institutional Repository. It has been accepted for inclusion in Open Access Theses and Dissertations by an authorized administrator of HKBU Institutional Repository. For more information, please contact repository@hkbu.edu.hk.
Understanding the Performance of Healthcare Services: A Data-Driven Complex Systems Modeling Approach

TAO Li

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

Principal Supervisor: Professor LIU Jiming

Hong Kong Baptist University

February 2014
Declaration

I declare that this thesis has been composed by myself under the guidance of my principal supervisor, Prof. LIU Jiming, and co-supervisor, Dr. CHEN Li, after registration for the degree of PhD at Hong Kong Baptist University. The thesis has not been previously included in any thesis, dissertation or report submitted to any institution for a degree, diploma or other qualification. All the reported results have been generated and verified using the methods as described. All sources of information have been acknowledged by means of references to the relevant publications.

Signature:____________________

Date: February 2014
Abstract

Healthcare is of critical importance in maintaining people’s health and wellness. It has attracted policy makers, researchers, and practitioners around the world to find better ways to improve the performance of healthcare services. One of the key indicators for assessing that performance is to show how accessible and timely the services will be to specific groups of people in distinct geographic locations and in different seasons, which is commonly reflected in the so-called wait times of services. Wait times involve multiple related impact factors, called predictors, such as demographic characteristics, service capacities, and human behaviors. Some impact factors, especially individuals’ behaviors, may have mutual interactions, which can lead to tempo-spatial patterns in wait times at a systems level. The goal of this thesis is to gain a systematic understanding of healthcare services by investigating the causes and corresponding dynamics of wait times.

This thesis presents a data-driven complex systems modeling approach to investigating the causes of tempo-spatial patterns in wait times from a self-organizing perspective. As the predictors of wait times may have direct, indirect, and/or moderating effects, referred to as complex effects, a Structural Equation Modeling (SEM)-based analysis method is proposed to discover the complex effects from aggregated data. Existing regression-based analysis techniques are only able to reveal pairwise relationships between observed variables, whereas this method allows us to explore the complex effects of observed and/or unobserved (latent) predictors on wait times simultaneously.

This thesis then considers how to estimate the variations in wait times with
respect to changes in specific predictors and their revealed complex effects. An integrated projection method using the SEM-based analysis, projection, and a queuing model analysis is developed. Unlike existing studies that either make projections based primarily on pairwise relationships between variables, or queuing model-based discrete event simulations, the proposed method enables us to make a more comprehensive estimate by taking into account the complex effects exerted by multiple observed and latent predictors, and thus gain insights into the variations in the estimated wait times over time.

This thesis further presents a method for designing and evaluating service management strategies to improve wait times, which are determined by service management behaviors. Our proposed strategy for allocating time blocks in operating rooms (ORs) incorporates historical feedback information about ORs and can adapt to the unpredictable changes in patient arrivals and hence shorten wait times. Existing time block allocations are somewhat ad hoc and are based primarily on the allocations in previous years, and thus result in inefficient use of service resources.

Finally, this thesis proposes a behavior-based autonomy-oriented modeling method for modeling and characterizing the emergent tempo-spatial patterns at a systems level by taking into account the underlying individuals' behaviors with respect to various impact factors. This method uses multi-agent Autonomy-Oriented Computing (AOC), a computational modeling and problem-solving paradigm with a special focus on addressing the issues of self-organization and interactivity, to model heterogeneous individuals (entities), autonomous behaviors, and the mutual interactions between entities and certain impact factors. The proposed method therefore eliminates to a large extent the strong assumptions that are used to define the stochastic properties of patient arrivals and services in stochastic modeling methods (e.g., the queuing model and discrete event simulation), and those of fixed relationships between entities that are held by system dynamics methods. The method is also more practical than
agent-based modeling (ABM) for discovering the underlying mechanisms for emergent patterns, as AOC provides a general principle for explicitly stating what fundamental behaviors of and interactions between entities should be modeled.

To demonstrate the effectiveness of the proposed systematic approach to understanding the dynamics and relevant patterns of wait times in specific healthcare service systems, we conduct a series of studies focusing on the cardiac care services in Ontario, Canada. Based on aggregated data that describe the services from 2004 to 2007, we use the SEM-based analysis method to (1) investigate the direct and moderating effects that specific demand factors, in terms of certain geodemographic profiles, exert on patient arrivals, which indirectly affect wait times; and (2) examine the effects of these factors (e.g., patient arrivals, physician supply, OR capacity, and wait times) on the wait times in subsequent units in a hospital. We present the effectiveness of integrated projection in estimating the regional changes in service utilization and wait times in cardiac surgery services in 2010-2011. We propose an adaptive OR time block allocation strategy and evaluate its performance based on a queuing model derived from the general perioperative practice. Finally, we demonstrate how to use the behavior-based autonomy-oriented modeling method to model and simulate the cardiac care system. We find that patients’ hospital selection behavior, hospitals’ service adjusting behavior, and their interactions via wait times may account for the emergent tempo-spatial patterns that are observed in the real-world cardiac care system.

In summary, this thesis emphasizes the development of a data-driven complex systems modeling approach for understanding wait time dynamics in a healthcare service system. This approach will provide policy makers, researchers, and practitioners with a practically useful method for estimating the changes in wait times in various “what-if” scenarios, and will support the design and evaluation of resource allocation strategies for better wait times management. By addressing the problem of characterizing emergent tempo-spatial wait time patterns in the cardiac
care system from a self-organizing perspective, we have provided a potentially effective means for investigating various self-organized patterns in complex healthcare systems.

**Keywords:** Complex Healthcare Service Systems, Wait Times, Data-Driven Complex Systems Modeling, Autonomy-Oriented Computing (AOC), Cardiac Care
Acknowledgements

First of all, there are no words to express the deepest gratitude I have to my principal supervisor, Prof. Jiming Liu. Throughout the past rewarding four years, I have deeply appreciated his guidance, support, patience, and encouragement. His enthusiasm for meaningful researches, serious attitude, excelsior spirit, profound knowledge, and efficient working method, will be beneficial to me throughout my life.

I wish to thank my co-supervisor, Dr. Li Chen, as well as Dr. Bo Xiao, Prof. Ho Cheung Leung, Dr. Kwok Wai Cheung, Prof. Yiu Ming Cheung, Dr. Yu Zhang, and Dr. Fion Lee in our department, for their help, advice, and encouragement on my research during the past years. I also wish to thank Prof. Eugene Santos, Jr. at Dartmouth College (USA), Prof. Toru Ishida at Kyoto University (Japan), Prof. Michael Carter at University of Toronto (Canada), Prof. Bo Yang at Jilin University (China) for the valuable suggestions and encouragement to me. I sincerely thank the external examiners, Prof. Joseph Tan and Prof. Qing Li, for valuable comments and suggestions on the previous version of the thesis.

I would also like to acknowledge the past and present members in our AOC research group, they are, Dr. Benyun Shi, Mr. Shang Xia, Dr. Chao Gao, Dr. Zhiwen Yu, Mr. Xiaofei Yang, Dr. Lei Fan, Ms. Yuan Bai, Ms. Hongjun Qiu, Dr. Shuaiqiang Wang, and Mr. Lailei Huang. I greatly appreciate their help and kindness to me.

I want to thank our departmental staff, Mrs. Letty Kwok, Ms. Suan Choi, and Ms. Kristina Tsang. They helped me deal with lots of study related matters. I also
want to give my thanks to the technicians in our department, Ms. Leah Poon, Mr. S. Y. Leung, and Mr. Johnson Chow for their technical support.

I acknowledge the University Grants Committee (UGC) and the Research Grants Council (RGC) of Hong Kong as well as the Department of Computer Science at HKBU for the financial support provided to me throughout my study.

My sincerely thank my lovely and kindly roommates, Ms. Jue Wang, Ms. Xuanzi Cui, Ms. Huiming Li, Dr. Xiuhua Ni, for their encouragement and kindness to me.

Last, but not the least, I would like to give my heartfelt thanks to my husband, Liang Luo, and my parents for their endless love as well as spiritual support to me in my life.
Abbreviations

ABM Agent-based modeling
AOC Autonomy-Oriented Computing
AOC-CSS AOC-based cardiac surgery service
CCN Cardiac Care Network of Ontario
CS-OR Cardiac surgery operating room
CU Catheterization unit
GP General practitioner
HHSC Hamilton Health Science Centre
LHIN Local Health Integration Network
KL Kullback-Leibler
LV Latent variable
MSMQ-EC Multi-server multi-queue with an entrance control queuing model
MV Measurement variable
NE North East
OPHRDC Ontario Physician Human Resources Data Center
OR Operating room
PCA Principle component analysis
PLS Partial least squares
RI Recent immigrant
SEM Structural Equation Modeling
SU Cardiac surgery unit
TC Toronto Central
# Table of Contents

Abstract ii  
Acknowledgements vi  
Abbreviations viii  
Table of Contents ix  
List of Tables xiv  
List of Figures xv  

Chapter 1 Introduction 1  
1.1 Background 2  
1.2 Motivations and Objectives 8  
1.2.1 Discovering the Direct, Indirect, and Moderating Effects of Observed and Latent Factors 8  
1.2.2 Estimating the Changes in Wait Times with Demographic Shifts 9  
1.2.3 Designing and Evaluating Service Management Strategies 11  
1.2.4 Characterizing Tempo-Spatial Patterns in Wait Times 13  
1.3 Contributions and Significance 16  
1.4 Structure of the Thesis 21  

Chapter 2 Literature Review 23  
2.1 Healthcare Service Systems 23
4.2.2 The Conceptual Model ......................................... 73

4.3 SEM Tests and Results ............................................. 74
4.3.1 Service Administrative Data in Ontario ....................... 74
4.3.2 SEM Tests ...................................................... 78
4.3.3 Test Results .................................................... 79

4.4 Discussion .......................................................... 83

4.5 Summary ............................................................ 86

Chapter 5  Projecting the Changes of Service Performance 89
5.1 Introduction ........................................................ 90

5.2 Integrated Projection ............................................. 93
5.2.1 SEM-Based Analysis ........................................... 94
5.2.2 Projection ....................................................... 95
5.2.3 Queueing Model Simulation ................................. 96

5.3 Estimating the Performance of Cardiac Surgery Services ........ 98
5.3.1 Demographic and Service Administrative Data in Ontario . 98
5.3.2 Relationships between Demographic Factors and Service
Characteristics ......................................................... 99
5.3.3 Service Performance Projection ............................... 101
5.3.4 The MSMQ-EC Queueing Model ............................. 102
5.3.5 Projection Results ............................................. 103

5.4 Summary ............................................................ 107

Chapter 6  Designing and Evaluating an Adaptive Strategy for Service
Management ......................................................... 108

6.1 Introduction ........................................................ 109

6.2 Designing an Adaptive OR Time Block Allocation Strategy ... 111

6.3 Modeling OR Services ............................................. 113

6.4 Simulation-Based Experiments ................................... 115
6.4.1 Experimental Settings ......................................... 115
Chapter 7  Characterizing the Tempo-Spatial Patterns in Patient
Arrivals and Wait Times 123

7.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 124
7.2 Empirical Tempo-Spatial Patterns in Cardiac Surgery Services . . . 128
  7.2.1 Aggregated Data . . . . . . . . . . . . . . . . . . . . . . . . . . 129
  7.2.2 Statistical Regularities . . . . . . . . . . . . . . . . . . . . . . 130
  7.2.3 Spatial Patterns . . . . . . . . . . . . . . . . . . . . . . . . . . 132
  7.2.4 Temporal Patterns . . . . . . . . . . . . . . . . . . . . . . . . . 136
7.3 AOC-CSS Modeling . . . . . . . . . . . . . . . . . . . . . . . . . . . 139
  7.3.1 Identifying Key Elements in Modeling . . . . . . . . . . . . . . 139
  7.3.2 Modeling Environment . . . . . . . . . . . . . . . . . . . . . . . 142
  7.3.3 Modeling Entities . . . . . . . . . . . . . . . . . . . . . . . . . . 144
  7.3.4 Designing Behavioral Rules . . . . . . . . . . . . . . . . . . . . 147
7.4 Simulation-Based Experiments . . . . . . . . . . . . . . . . . . . . . 149
  7.4.1 Experimental Settings . . . . . . . . . . . . . . . . . . . . . . . 150
  7.4.2 Statistical Regularities in Patient Arrivals and Wait Times . . . 152
  7.4.3 Patient-Attraction and Patient-Distribution Degrees of LHINs 154
  7.4.4 Tempo-Spatial Patterns in Patient Arrivals and Wait Times . . . 155
7.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 156
  7.5.1 Explaining the Underlying Causes of Tempo-Spatial Patterns . 156
  7.5.2 Sensitivity Analysis . . . . . . . . . . . . . . . . . . . . . . . . . 163
7.6 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 167

Chapter 8  Conclusions and Future Work  168

8.1 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 168
8.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 172
List of Tables

3.1 The names, sizes, and scopes of LHINs in Ontario, Canada ............ 47
3.2 The measurement values for the geodemographic profiles of LHINs providing cardiac surgery services (2006) ........................................... 54
3.3 Cardiac surgery service utilization from 2004 to 2007 in Ontario hospitals ................................................................. 55
3.4 Summary statistics for the geodemographic factors and cardiac service utilization in Ontario between 2004 and 2007 ......................... 56
3.5 Hypothesis testing results .......................................................... 59
4.1 The relationship between the CCN member hospitals and the LHINs 76
4.2 Cardiac surgery statistics from January 2008 to March 2008, obtained from the CCN ............................................................... 77
4.3 A summary of the characteristics of the CU and SU ...................... 78
4.4 Hypothesis testing results .......................................................... 83
5.1 The training data set ................................................................. 100
5.2 The estimated values for cardiac surgery utilization and performance (average value in a month) ........................................ 104
6.1 Cardiac surgery services in the HHSC in 2004 ............................ 116
7.1 The p-values of power-law tests for distributions of absolute wait time variations with respect to different $P_r$ ........................................... 165
List of Figures

1.1 A conceptual framework of a healthcare service system. .................. 3
1.2 A summary of the major contributions of the thesis. ....................... 16

2.1 A conceptual model illustrating a healthcare service system. ............. 24

3.1 A schematic diagram illustrating the use of an SEM-based analysis to explore the complex effects of certain geodemographic profiles on healthcare service utilization. (a) The research focus of this chapter (highlighted in red) is part of the larger context of understanding a healthcare service system. (b) Exploring underlying effects using an SEM-based analysis. ................................. 44
3.2 Conceptual model for this study. +/-: a positive/negative relationship between two LVs. ............................... 51
3.3 Population distribution across the selected cities and towns in Ontario. The populations of cities and towns in Ontario follow a power-law distribution (correlation coefficient $R = -0.922$, standard deviation $SD = 0.2441$, $p < 0.0001$). The selected cities and towns have populations larger than 40,000 and cover approximately 90.72% of Ontario’s population. ................................. 52
3.4 SEM test results: the effects of population size and age profile on service utilization. ........................................ 57
3.5 SEM test results: service accessibility as a moderator. ................. 58
3.6 SEM test results: educational profile as a moderator. ................. 59
4.1 A schematic diagram illustrating the use of an SEM-based analysis to examine the effect of a preceding unit’s characteristics on the wait times of a subsequent unit. (a) The research focus of this chapter (highlighted in red color) is part of the larger context of understanding a healthcare service system. (b) Exploring underlying effects using an SEM-based analysis.

4.2 The unit framework for cardiac care, drawn from the cardiac treatment guidelines. ECG: Electrocardiogram; PTCA: Percutaneous Transluminal Coronary Angioplasty; PCI: Percutaneous Coronary Intervention.

4.3 The research framework, summarizing the impact factors for throughput and wait times.

4.4 Conceptual model for this study. Cath: catheterization; Surgery: cardiac surgery; H1-H9: the research hypotheses; +/-: a positive or a negative relationship between two variables.

4.5 PLS test results based on a formative measurement model. Cath: catheterization; Surgery: cardiac surgery.

4.6 PLS test results for the extended two-layer wait time model with the low-risk profile in the SU. Cath: catheterization; Surgery: cardiac surgery.

4.7 PLS test results for the extended two-layer wait time model with the medium-risk profile in the SU. Cath: catheterization; Surgery: cardiac surgery.

4.8 PLS test results for the extended two-layer wait time model with the high-risk profile in the SU. Cath: catheterization; Surgery: cardiac surgery.
5.1 A schematic diagram illustrating the use of integrated projection to project the changes in service performance with respect to demographic shifts. (a) The research focus of this chapter (highlighted in red) is part of the larger context of understanding a healthcare service system. (b) Carrying out an estimation based on integrated projection. ............................... 90

5.2 A schematic diagram of the three-step integrated projection method and its application to cardiac surgery services. ......................... 92

5.3 The basic components in SEM. ........................................ 94

5.4 An schematic diagram of a general queueing model for healthcare. E: elective; S: semi-urgent; U: urgent. ................................. 97

5.5 The hypothetical relationships between the demographic profiles, service utilization, and service performance. ......................... 101

5.6 The PLS test results. (a) Correlations between the LVs (b) Path coefficients between the LVs. RI: recent immigrant. ...................... 105

5.7 The estimated service utilization and resulting simulated queue lengths from 2010 to 2011. ........................................ 106

5.8 The estimated service utilization and resulting simulated average wait times from 2010 to 2011. ................................. 107

6.1 A schematic diagram illustrating the design and evaluation of an adaptive strategy for improving time block allocations in ORs. (a) The research focus of this chapter (highlighted in red) is part of the larger context of understanding a healthcare service system. (b) Designing and evaluating a service management strategy for OR time block allocation. ................................. 109

6.2 The OR scheduler with a feedback mechanism. ...................... 112

6.3 The mechanism for updating the OR time blocks for urgent surgeries. 113

6.4 A multi-priority, multi-server, non-preemptive queueing model with an entrance control mechanism. ................................. 114
6.5 The simulated queue lengths over three years (Inserted plots: the time blocks for urgent surgeries allocated with the adaptive strategy; \( \delta_0 = 7 \) per week, \( T = 4 \) weeks, \( \theta_1 = 2 \times T \), \( \theta_2 = 1 \times T \), \( \Delta p = 1 \), \( \Delta q = 1 \)).

6.6 OR utilization with respect to different initial urgent OR time blocks (AS: adaptive strategy; BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; \( \delta_0 = 7 \) per week, \( T = 4 \) weeks, \( \theta_1 = 2 \times T \), \( \theta_2 = 1 \times T \), \( \Delta p =, \Delta q = 1 \)).

6.7 The trade-offs (i.e., BNS/UUB) of the adaptive strategy with respect to different \( T \) (BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; \( \delta_0 = 7 \) per week, \( \theta_1 = 2 \times T \), \( \theta_2 = 1 \times T \), \( \Delta p =, \Delta q = 1 \)).

6.8 OR utilization with respect to adjustment thresholds (BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; \( \delta_0 = 7 \) per week, \( T = 4 \) weeks, \( \theta_1 = 2 \times T \), \( \theta_2 = 1 \times T \), \( \Delta p =, \Delta q = 1 \)).

6.9 OR utilization with respect to step sizes (BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; \( \delta_0 = 7 \) per week, \( T = 4 \) weeks, \( \theta_1 = 2 \times T \), \( \theta_2 = T \)).

6.10 Queue lengths generated by different block allocation strategies (Setting I: \( \delta_0 = 7 \) per week, \( T = 4 \) weeks, \( \theta_1 = 2 \times T \), \( \theta_2 = 1 \times T \), \( \Delta p =, \Delta q = 1 \); Setting II: \( \delta_0 = 5 \) per week, \( T = 4 \) weeks, \( \theta_1 = 2 \times T \), \( \theta_2 = 1 \times T \), \( \Delta p =, \Delta q = 1 \)).

7.1 A schematic diagram illustrating the use of behavior-based autonomy-oriented modeling to characterize the emergent tempo-spatial patterns in patient arrivals and wait times in a healthcare service system. (a) The research focus of this chapter (highlighted in red) is part of the larger context of understanding a healthcare service system. (b) Modeling, simulating, and analyzing the behavior of a healthcare service system.
7.2 The complex cardiac surgery system in Ontario, Canada. The illustrated tempo-spatial patterns are observed from the secondary data on cardiac surgery services between January 2005 and December 2006. LHIN: Local Health Integration Network; H: hospital.127

7.3 The distribution of hospitals that provide cardiac surgery services across the LHINs in Ontario, Canada (adapted from the Ontario LHINs map given at http://www.lhins.on.ca). Red dots denote hospital locations and the numbers correspond to the LHIN IDs, following Table 3.3.129

7.4 The statistical distribution of variations in patient arrivals for cardiac surgery services in Ontario, Canada, between January 2005 and December 2006. The normality of the distribution passes the Kolmogorov-Smirnov test. Standard deviation $SD = 0.226$.131

7.5 The statistical distribution of the absolute variations in median wait times for cardiac surgery services in Ontario, Canada, between January 2005 and December 2006. The distribution follows a power law with a power of $-1.36$. (power-law test based on Clauset’s method [215]: $p < 0.1$; linear fitness (red line): $p < 0.001$; standard deviation $SD = 0.28$).132

7.6 The distribution of cardiac surgery patients with respect to their LHIN residence between 2007 and 2008 in Ontario, Canada.133

7.7 The distribution of LHINs’ patient-attraction and patient-distribution degrees.135

7.8 The changes in patient arrivals for cardiac surgery services in (a) Ontario and (b) five hospitals that provide cardiac surgery services.137

7.9 The changes in the median wait times for cardiac surgery services in (a) Ontario and (b) five hospitals that provide cardiac surgery services.138
7.10 The effects of impact factors on patient-GP mutual decisions on hospital selection and the local feedback loops. “+”: positive relationship between two factors; “−”: negative relationship between two factors. ................................................................. 142

7.11 The number of patient arrivals versus the number of treated cases in the cardiac surgery service in Ontario, Canada, between January 2005 and December 2006. ................................................................. 150

7.12 Distributions of the variations in simulated and observed patient arrivals in cardiac surgery services. SD: standard deviation. .................. 152

7.13 Distribution of simulated absolute wait time variations (by month) in cardiac surgery services. The distribution follows a power law with a power of -1.47 (power-law test based on Clauset’s method [215]: \( p < 0.1 \); linear fitness (red line): \( p < 0.0001 \); standard deviation \( SD = 0.183 \)). ................................................................. 153

7.14 Distributions of simulated and observed wait time variations in the cardiac surgery service. ................................................................. 154

7.15 Comparison of the simulated and observed distributions of LHINs’ patient-attraction and patient-distribution degrees. ......................... 155

7.16 The observed and simulated temporal patterns in patient arrivals in hospitals H1-H6 ................................................................. 157

7.17 The observed and simulated temporal patterns in patient arrivals in hospitals H7-H11. ................................................................. 158

7.18 The observed and simulated temporal patterns in wait times in hospitals H1-H6 ................................................................. 159

7.19 The observed and simulated temporal patterns in wait times in hospitals H7-H11. ................................................................. 160
7.20 The dynamically changing preferences of patients residing in the city of Brampton (in LHIN 5) to the four neighboring hospitals. (a): H4, Trillium Health Centre; (b): H5, St. Michael’s Hospital; (c): H6, Sunnybrook Hospital; (d): H7, University Health Network. The shaded areas in this figure represent the warm seasons in Ontario, Canada. ................................................................. 162

7.21 Distribution of simulated absolute wait time variations (calculated by week) in cardiac surgery services. The distribution follows a power law with a power of -2.19 (power-law test based on Clauset’s method [215]: $p < 0.1$; linear fitness (red line): $p < 0.0001$; standard deviation $SD = 0.331$). ................................................................. 164

7.22 Distribution of simulated absolute wait time variations (calculated by half-month) in cardiac surgery services. The distribution follows a power law with a power of -1.86 (power-law test based on Clauset’s method [215]: $p < 0.1$; linear fitness (red line): $p < 0.001$; standard deviation $SD = 0.38$). ................................................................. 165

7.23 Distributions of absolute wait time variations (by month) in cardiac surgery services with respect to different $P_r$. ................................................................. 166
Chapter 1

Introduction

In this chapter, we introduce the background to this thesis, which is the challenges to understanding a healthcare service system with the aim of improving service performance through wait times. We then outline the motivation for and research questions describing the study. This thesis proposes a systematic approach for investigating the causes of the dynamics of a specific healthcare service system and relevant patterns in its wait times, from four perspectives.

1. Exploring the effects of multiple impact factors on patient arrivals and wait times;

2. Estimating the changes in wait times with respect to demographic shifts;

3. Designing and evaluating service management strategies for better use of service resources and improved wait times;

4. Characterizing the tempo-spatial patterns of wait times by taking into account patients’ hospital selection behavior with respect to certain impact factors.

We present the major contributions of this thesis, which cover computational methods and real-world applications to address the four questions. Finally, we introduce the overall structure of the thesis.
1.1 Background

Healthcare is of great importance in promoting the public health and wellness of people [1, p.2]. *Wait time*, how long patients wait for desirable services, is a measure of a healthcare service’s performance that has received significant attention. Most waits are unnecessary [2] and cause patient deterioration [3]. Improving wait times has recently been at the forefront of healthcare planning and management in many countries [2, 4]. However, there are still long wait times [5], and high variations in wait times in different locations and at different times of day, which may follow *tempo-spatial patterns* [6]. We therefore need a better understanding on the causes of the dynamics and of the relevant patterns in wait times in a specific healthcare service system, which is the goal of this thesis.

Figure 1.1 presents a conceptual framework of the healthcare service system that is investigated here. The healthcare service system can be divided into a system that includes multiple service providers that deliver healthcare services (which are hospitals here), located in different cities or regions; (2) the system’s input, including patients receiving services; and (3) the system’s output, which involves multiple indicators representing the performance of each healthcare service. As shown in the middle box of this figure, each service provider’s *profiles* [7, p.9] differ in terms of the number and types of personnel, such as physicians and general practitioners (GPs); the on-site facilities, such as operating rooms (ORs) and laboratories; the included organizations, such as units to manage personnel and facilities for providing services to patients [8]; and its cost and financing. Service providers also vary in their *service management behavior*, such as scheduling time blocks in ORs [9, 10] and referring patients to other service providers [11, 12]. A service provider’s profiles and service management behavior, called its *supply factors* [13], determine the actual delivery of services to patients to a significant extent, and thus directly affect the variations in wait times.
Figure 1.1: A conceptual framework of a healthcare service system.
The left-hand-side box in Figure 1.1 shows a healthcare service system’s input, the individuals residing in a specific region, which is referred to as a population here. The population in a region may exhibit distinct profiles because individuals have differences in their ages [14], education levels [15, 16], social networks [17, 18], and other personal characteristics. The profiles of a population also refer to the built and natural environment, such as geographic access to services [19, 20] and seasonal weather [21, 22]. A population’s profiles affect the incidence of disease, which directly determines to a large extent the number of patient arrivals at healthcare services. Autonomous service utilization behavior refers to an individual’s behavior that is governed by their decision making regarding the selection and use of specific service providers, and determines whether and which service provider an individual will use. A population’s profiles and behavior, called the demand factors [23], may result in patient arrivals varying at different times [22] and in different regions [24].

The right-hand box in Figure 1.1 illustrates the system output considered in this thesis, which is service performance represented by various indicators, such as wait times, throughput, in-hospital mortality rate, and re-admission rate. The indicators are categorized into eight dimensions [25] to measure the service performance using measures such as the timely accessibility of services; efficiency in dealing with patients; the quality, safety and appropriate use of services; and patients’ experience with the received services. It should be noted that most of the indicators and some of the impact factors are abstract concepts (also called latent variables, (LVs)) that should be estimated using several observed variables or measurements. For instance, the indicator wait times can be judged by median wait time, ninetieth percentile wait time, or queue length [5, 24], each of which represents one aspect of the wait times.

Demand and supply factors can exert direct, indirect or moderating effects on patient arrivals and/or wait times. Direct effects measure how a dependent variable changes when a predictor variable increases or decreases by one unit. The direct effect is commonly represented by a positive or a negative path coefficient (weight)
in statistics. For instance, studies have found that the factor of service capacity may directly influence the variation in wait times [26, 27]. Service capacity therefore has a direct effect on wait times. *Indirect effects* denote a predictor variable that influences a dependent variable through a third variable. For instance, the factor population size imposes an indirect effect on wait times, as a larger population may translate into a greater number of patient arrivals [28, p.59], which is one of the direct causes of wait times [29]. *Moderating effects* measure how a third independent variable may change the direction and/or the strength of the relationship between a predictor variable and wait times. For instance, the prevalence of smoking and inactivity, two traditional cardiovascular risk factors, in the less-educated population [15, 22] suggests that a higher proportion of well-educated individuals in the population may mitigate the pressure of population size on patient arrivals, and thus ease the burden on wait times [30]. In this thesis, the direct, indirect, and moderating effects are referred to as *complex effects*. One of the major challenges in gaining a qualitative understanding of the dynamics of wait times is the complex effects of multiple impact factors on wait times.

Discovering the effects of factors may benefit from the use of various types of aggregated data that is publicly available, and thus is easier to access than first-hand medical data such as electronic health records. Two types of aggregated data [31, 32] summarizing healthcare service systems and how people are served have been collected recently, to monitor the operations and performance of healthcare service systems. *Survey data* [32] collected and published by government agencies, stakeholders in healthcare, and service providers can provide information about the profiles and behavior of individuals and service providers. For instance, the Census Bureau provides survey data about the demographics, socioeconomics, and land of a region [33]. Healthcare organizations such as the Cardiac Care Network of Ontario (CCN)\(^1\) and the Ontario Physician Human Resources Data Center (OPHRDC)\(^2\) have published survey data on patient hospital selection behavior [34] and the number of

\(^1\)http://www.ccn.on.ca/ccn\_public/FormsHome/HomePage.aspx
\(^2\)https://www.ophrdc.org/Home.aspx
physicians in each hospital in Ontario by specialty. Administrative data collected by healthcare organizations and service providers “for administrative, regulatory, healthcare operations...purposes” [31, p.73] can represent information about service profiles, behavior, and performance. For instance, ICES Atlas\(^3\) and the CCN have reported statistical data to show the performance of cardiac care services in Ontario in terms of the throughput, median wait time, ninetieth percentile wait time, and queue length [24, 35].

These data hold the potential for understanding what causes the tempo-spatial patterns in wait times in a healthcare service system, using the following five approaches.

1. **Discovering the complex effects of certain population profiles on service utilization from aggravated data.** Demographics, socioeconomic status, and other population profiles of a region are recognized as important determinants of healthcare service utilization [36, 37]. Here, service utilization is defined as the number of patient arrivals, which indirectly affects wait times in healthcare services. However, few studies have explored how specific geodemographic factors indirectly and moderately affect service utilization. Thus, identifying the complex effects of certain population profiles on service utilization may fill this gap in the literature and suggest applicable implications.

2. **Exploring the complex effects of one unit’s characteristics on the wait times of a subsequent unit.** Service utilization and capacity have been shown to significantly affect the throughput and wait times within a unit in a hospital. However, as most of the units in a hospital are networked via patient flow, it is doubtful whether the characteristics of one unit, such as service utilization, capacity, throughput, and wait times, would affect the wait times of subsequent units. There is usually a “funnel and filter effect” [38, p.163] between two

---

\(^3\)http://www.ices.on.ca/webpage.cfm?site_id=1&org_id=67&morg_id=0&gsec_id=0&item_id=3559&type=atlas
temporally related units. That means, preceding units determine “the absolute number and speed of the throughput for patients proceeding to the subsequent units” [38, p.163]. Thus, we must investigate how the wait times of a unit are associated with the characteristics of its preceding units, which will suggest how wait time management can be improved through organizational structure [39].

3. *Estimating the changes in wait times with respect to demographic shifts.* Once the relationships between demand factors, supply factors, service utilization, and wait times are identified, it is possible to make a projection of future wait times with respect to variations in certain factors. As demographics in many areas of the world are experiencing notable changes due to population growth and aging, healthcare service providers may face more patient arrivals, potentially burdening existing healthcare services. Estimating the changes in service utilization and wait times in accordance with demographic shifts will therefore help for healthcare administrators to improve the planning and allocation of healthcare resources.

4. *Designing and evaluating service management strategies for improving wait times.* In addition to the associations between a service provider’s profiles and wait times, service management behavior also plays a significant role in service delivery and wait times. We therefore need to design new service management strategies to improve wait times. For instance, current practices in allocating time blocks in ORs are somewhat ad hoc, based primarily on the allocations made in previous years. Allocating ORs with such a relatively static strategy may not effectively use the available resources, due to unpredictable patient arrivals. New strategies for allocating time blocks in ORs and evaluating the effectiveness of the strategies for better use of ORs and improved wait times are therefore needed.

5. *Explaining tempo-spatial patterns in wait times.* Patients autonomously
decide whether or not to go to a hospital based on specific factors, such as the distance between home and the hospital, service profiles, and service performance. Patients’ autonomous behavior may result in dynamically changing patient arrivals at hospitals, leading to variations in wait times, which may follow tempo-spatial patterns. We will therefore model and characterize these emergent tempo-spatial wait time patterns using individuals’ autonomous behavior with respect to various impact factors, which will offer crucial insights into the nature of healthcare service systems.

1.2 Motivations and Objectives

In this section, we present the motivations for systematically understanding healthcare services in terms of wait times. We present the four specific objectives of the study, to explore the complex effects of multiple factors on patient arrivals and wait times; to estimate the changes in wait times with demographic shifts; to propose and evaluate a new service management strategy for improving wait times; and to characterize the tempo-spatial patterns of wait times by modeling patients’ autonomous behavior.

1.2.1 Discovering the Direct, Indirect, and Moderating Effects of Observed and Latent Factors

From a computational perspective, discovering the direct, indirect, and/or moderating effects of certain population profiles on service utilization, and of one unit’s characteristics on the wait times of a subsequent unit, can be transformed into one research question: how to explore the complex relationships of predictor variables (e.g., impact factors) with dependent variables (e.g., service utilization and wait times). Two specific research issues must be addressed to answer this question.

1. Modeling LVs: As the concerned variables include those that cannot be directly
observed, how can we build a mathematical model to infer an LV from other observed variables?

2. **Modeling complex relationships between variables**: How can we model concurrently direct, indirect, and moderating relationships between observed variables and/or LV using mathematical equations? How can we infer *path weights* from aggregated data to represent these relationships.

Studies have usually relied on traditional multivariate statistical methods to discover the relationships (i.e., path weights) between variables from limited aggregated data, such as regression [40, 41]. These methods usually model the relationship between the dependent and independent variables as a linear, logistic, or other type of function [42]. However, these methods are limited when constructing LVs and modeling the complex relationships between variables, rather than the pairwise relationships between dependent and independent variables.

*Structural Equation Modeling* (SEM) provides a means for addressing the two research issues. In general, SEM has the ability to construct LVs [43] and permits the exploration and confirmation of complex relationships between variables concurrently [43, 44].

The first objective of this thesis is therefore *to propose a method using SEM to investigate the direct, indirect, and/or moderating relationships between the observed variables and LVs based on aggregated data.*

### 1.2.2 Estimating the Changes in Wait Times with Demographic Shifts

Estimating the changes in wait times with demographic shifts is, in essence, a problem of estimating the dynamics of specific dependent variables with respect to the variations in specific predictor variables. Three specific research issues must be addressed to answer this question.

1. **Exploring complex relationships between variables**: Given aggregated data that
describes the dependent and predictor variables of interest, how can we explore the complex relationships between these variables?

2. Estimating the changes in specific dependent variables: Given the trends in the changes of certain predictor variables, can we obtain the variations in specific dependent variables in the future, based on the identified relationships between the variables?

3. Presenting the dynamics of estimated variables: As the estimation results based on variable relationships are somewhat sketchy, how can we determine the dynamically changing process of focal dependent variables over time?

Existing studies dealing with these research issues in the healthcare context can be classified into three categories.

1. Estimation based on pairwise variable relationships. These studies have aimed to discover the effects of predictors on specific dependent variables using traditional statistical methods, such as regression, which can be used to forecast the changes in dependent variables if the predictors vary. For instance, the demographic profiles of population age and ethnicity are two of the most important determinants for cardiac surgery service utilization. Cardiovascular risk factors, such as diabetes, hypertension, and obesity, are higher in the age group 50 years and above [14, 22] and vary in different ethnic groups [22, 45]. However, these commonly used traditional statistical methods may not be able to model LVs and reveal complex relationships between variables.

2. Forecast based on specific patterns of variables. These studies have focused on finding the underlying patterns in certain variables, which can then be used to make scenario-based predictions. For example, previous research has found that there are more patient arrivals in the winter compared to other seasons, because of cold weather [22]. The arrivals also vary depending on
the time of day and the day of the week [46, 47]. Some studies have tried to make forecasts for specific scenarios, such as predicting emergency department arrivals in a disaster [48] or during the pandemic influenza season [49]. However, predictions based on identified patterns require that the scenarios for finding patterns and those for predictions are the same. This requirement is difficult to satisfy in the situation considered in this thesis, as the scenario may change if specific predictor variables vary.

3. *Prediction based on modeling methods.* These studies have attempted to present the dynamically changing service performance of a healthcare service system in a given scenario. They have usually used queueing models [50] and discrete event simulations [51] to examine the dynamics of service utilization in a healthcare service system and to assess performance variations based on so-called “what-if” studies [50]. However, these studies have not addressed how service performance changes in response to demographic shifts.

Motivated by the above observations, the second objective of this thesis is to propose a method for estimating the changes in service utilization and performance in response to demographic shifts. The method will use the SEM technique to identify the complex relationships between variables; estimate healthcare service utilization based on the discovered variable relationships, which are assumed to hold during the time considered; and demonstrate the dynamics of estimated service utilization and service performance over time.

### 1.2.3 Designing and Evaluating Service Management Strategies

Two specific research questions must be answered to design and evaluate service management strategies to cope with unpredictable patient arrivals, with the goal of better healthcare resources use and improving wait times.

1. *Designing strategies:* In view of current healthcare service management
strategies and the dynamics of patient arrivals, what new strategies can we propose?

2. Evaluating strategies: With a given scenario describing the patterns of stochastic patient arrivals and the profiles of a specific service provider, how can we model service behavior to provide a test-bed for evaluating the effectiveness of new strategies?

Existing studies have often used mathematical methods, such as mathematical programming to optimize specific measurements, to design better service management strategies. The management of time blocks in ORs in cardiac surgery services can be used as an example. One of the common questions that are considered when designing new OR time block allocation strategies is how many OR time blocks should be reserved to cope with the unpredictable arrival of urgent patients. Reserving more time blocks than are needed can cause decreased OR use, a longer waiting list, and a longer waiting time for non-urgent surgeries. Reserving insufficient time blocks can increase the risk to urgent patients, incur more bumped non-urgent surgeries, and prolong wait times for those bumped cases. To improve the use of OR time blocks, existing studies have used mathematical methods (e.g., job shop scheduling models) to compute the optimal number of reserved urgent time blocks, aiming to maximize OR time block use while minimizing the overtime/cancellation of surgeries [52]. However, these studies have not considered that patient arrivals are dynamic because of the number of impact factors involved, such as the weather and patients’ service utilization behavior [22]. Therefore, the theoretical optimal solution may not perform well in an actual healthcare service.

One of the common ways to model service behavior to evaluate different service management strategies, is to simulate the service operations with different strategies and to compare the simulated results based on certain measurements. In healthcare service research, the queueing model is commonly used to characterize the behavior of a service provider and examine the dynamics of service performance, such as wait times and queue length. In general, the queueing model describes stochastic patient
arrivals and the services delivered by a healthcare service system as a continuous-time or a discrete-time Markov chain, where the system state corresponds to the number of patients in the system. The expected queue lengths and wait times can be mathematically analyzed and the dynamics of service performance can be simulated with a specific queueing method.

The third objective of this thesis is therefore, within the context of time block management in ORs, to design an adaptive service management strategy to improve the use of service resources with respect to unpredictable patient arrivals, and evaluate the effectiveness of the adaptive strategy in improving service performance.

1.2.4 Characterizing Tempo-Spatial Patterns in Wait Times

Explaining the tempo-spatial patterns in wait times can be defined as a modeling and simulation research question: how to model and simulate individuals’ behavior and interactions with respect to certain impact factors, to reveal the underlying mechanisms that account for the observed tempo-spatial patterns in wait times. Three specific issues must be addressed to answer this question.

1. **Modeling entities**: What specific autonomous entities potentially play significant roles in the emergent patterns and thus should be modeled, and how should they be defined within the model? Here, entity indicates individuals that can autonomously make decisions, such as patients and GPs, and conceptual components, such as hospitals.

2. **Modeling entities’ behavior, interactions, and local feedback loop(s) with respect to certain factors**: Which behavior by the entities, major impact factors, and interactions between the two, are relevant to the observed tempo-spatial patterns in wait times and hence should be investigated and modeled? As local feedback loops [53] formed by relationships between the variables may amplify or dampen the effects of factors or entities’ behavior
on service performance and thus result in nonlinear dynamics in the system, which local feedback loop(s) should therefore be modeled? How can we model entities’ behavior and formulate the rules that govern this behavior, while taking into account the identified effects of factors and the heterogeneity of entities?

3. Carrying out simulation-based experiments: What tempo-spatial patterns at a systems level emerge from the simulation? Are the simulated emergent patterns similar to those observed in the real world? If a simulation based on the developed model can reproduce the tempo-spatial patterns observed in the real world, what are the underlying mechanisms that account for the emergent patterns?

Existing studies have usually used stochastic modeling and simulations, system dynamics, and agent-based modeling (ABM) to model the behavior of a healthcare service system and simulate the dynamics or tempo-spatial patterns of wait times. We summarize the major characteristics of these methods below.

1. **Stochastic modeling and simulations** aim to model a service system by defining the service profiles, service management behavior, and stochastic properties, such as Poisson arrivals and exponential services. Two classic methods are the queueing model and discrete event simulation. As these methods require that assumptions be made regarding the stochastic properties, they may meet a number of difficulties in characterizing the tempo-spatial patterns of wait times. The assumptions about the stochastic properties of a healthcare service system may be strong and not always hold in the real world. Further, these methods cannot explore how entities’ behavior and interactions result in the emergent tempo-spatial patterns at a system level, because they do not aim to model entities’ heterogeneous behavior.

2. **System dynamics** models a focal system as a causal loop diagram [54]. A system dynamics model contains entities (referred to “stocks” in this
method) that accumulate or are exhausted over time, and their interactions (referred to “flows” in this method) are usually represented by first-order differential or integral equations [55]. Unlike other methods, system dynamics pays special attention to modeling the internal interactions between entities and local feedback loops [53] within a system. However, it is difficult to model the heterogenous and autonomous behaviors of each entity with this model, as it assumes that each entity’s behavior is fixed. This method therefore cannot explain how entities’ behavior and interactions cause the emergent patterns in a healthcare service system.

3. **ABM** [56] aims to model a healthcare service system by defining agents (i.e., entities), their behavior, and interactions. However, traditional ABM faces a major challenge in characterizing system-level emergent patterns: it lacks the general principles to explicitly indicate which fundamental behavior of and interactions between agents play crucial roles in the emerging tempo-spatial patterns and therefore should be modeled. Many existing models based on ABM are more or less ad hoc with a major focus of delicately defining agents, whereas few of them pay attention to explaining the underlying mechanisms for emergent patterns in a healthcare service system.

Unlike the more commonly used methods, **autonomy oriented computing** (AOC) [57] provides a potentially effective way to investigate the emergent patterns in wait times from a self-organizing systems perspective. In essence, “an AOC system is a multi-agent system,” but it differs from a traditional multi-agent system in that it aims to address “the issues of self-organization, self-organized computability, interactivity...” [58, p.9] in solving problems and modeling complex systems. Here, the notion of **self-organization** means that entities in a system may self-organize their behaviors through their mutual interactions, in accordance with the changes in multiple impact factors. **AOC-by-prototyping** [57] defines local-autonomy-oriented entities; defines the environment that the entities reside in and interact with; models entities’ autonomous behavior, interactions, and feedback loops, taking into account
the effects of multiple factors; and uncovers the essential mechanisms of positive-feedback-based aggregations or negative-feedback-based regulations at a systems level through model-based simulations.

Based on the above motivations, the fourth objective of this thesis is to propose a modeling method using AOC to uncover the working mechanisms that account for the emergent patterns in wait times in a specific healthcare service system.

1.3 Contributions and Significance

Figure 1.2 summarizes the major contributions of this thesis. As presented in the left box of Figure 1.2, this thesis presents a data-driven complex systems modeling approach using four specific methods to understand the causes of the dynamics of and relevant patterns in wait times in a healthcare service system. As highlighted in the right box of Figure 1.2, this thesis uses these methods to unveil potential reasons for and the working mechanisms behind the observed patterns in wait times in a real-world cardiac care system in Ontario, Canada. Further details on these contributions are presented below.
1. Proposing an *SEM-based analysis* method for exploring the complex relationships between demand factors, supply factors, service utilization, and/or wait times.

We propose an *SEM-based analysis* method to explore the complex relationships between observed variables and LVs based on aggregated data, as most of the existing studies on wait times in healthcare have found this objective difficult. This method explicitly follows four steps: developing hypotheses related to the direct, indirect, and moderating relationships between observed variables and LVs, based on the literature; constructing a conceptual model that contains the complex relationships between the observed variables and LVs under consideration; using SEM to test the hypotheses with aggregated data; and interpreting the test results by comparing the discovered relationships between variables with those reported in the literature or observed in the real world. Compared to existing methods, our proposed SEM-based method enables us to explore the direct, indirect, and moderating effects of certain observed or latent impact factors on service utilization and/or wait times in a healthcare system.

To demonstrate and implement the SEM-based analysis method, we explore the complex effects of certain demand or supply impact factors on patient arrivals and wait times in a real-world cardiac care system in Ontario, Canada. Chapter 3 presents the demand side of the cardiac care system. The direct and moderating effects of specific geodemographic profiles (i.e., the population size, age profile, service accessibility, and education profile) on service utilization of cardiac surgery services are examined. Based on publicly-available aggregated data on the geodemographic profiles of Ontario and the corresponding cardiac surgery services between 2004 and 2007, the data test results show that service accessibility and the education profile alleviate the effects of the population size and/or age profile on service utilization. This is a novel finding that, to the best of our knowledge, has not been previously reported. This finding reveals that the changes in population profiles due to population growth and aging may significantly affect the
use of cardiac surgery services. It also suggests the importance of considering the geodemographic profiles of a geographic area and, in some cases, its neighboring areas, when allocating healthcare service resources, thus strategically improving service utilization and reducing wait times.

Chapter 4 shows the analysis of the supply side of the cardiac care system. The effects of a catheterization unit’s (CU’s) characteristics (e.g., service utilization, capacity, throughput, and wait times) on wait times in the subsequent cardiac surgery unit (SU) are investigated. Based on published aggregated data on catheterization and cardiac surgery services, our findings show that the wait times in the CU have a direct positive effect on the wait times in the SU. This is a novel result, as prior research has seldom examined the influence of one unit’s wait times on the wait times in a subsequent unit in the patient flow process. Our findings also show that the service utilization and wait times of a preceding unit are good predictors for the wait times of subsequent units, suggesting that such cross-unit effects must be considered if the wait times in a healthcare service system are to be alleviated.

2. Proposing an integrated projection method for estimating the changes in service utilization and service performance with respect to demographic shifts.

To further use the identified variable relationships, we propose an integrated projection method to estimate how service utilization and service performance in a healthcare service system change with respect to demographic shifts. Our proposed projection method consists of three steps: applying SEM to identify complex relationships between demographic profiles and healthcare service characteristics (e.g., service utilization, physician supply, OR capacity, throughput, and wait times); estimating the changes in service utilization and performance using the discovered variable relationships, which are assumed to hold during the estimation period; and constructing specific queueing models and conducting simulation-based experiments to present the dynamics of the estimated service performance over time. Compared to existing methods for estimating service utilization and service performance of a healthcare service system, our method is
able to make a comprehensive projection of service utilization and performance, taking into account the complex effects exerted by multiple observed and latent factors, and to demonstrate the temporal changes in predicted service utilization and performance by means of a queueing model simulation.

Chapter 5 implements this method to estimate the regional use of cardiac surgery services in Ontario between 2010 and 2011, based on statistics between 2005 and 2007. The findings show that our analytical method is able to identify the complex effects of the age profile, recent immigrant profile, and the characteristics of cardiac surgery services on service utilization; describe the variations in service utilization with respect to demographic shifts; and demonstrate the temporal changes in estimated cardiac surgery performance in terms of queue length. The work presented in this chapter can help a healthcare service system to dynamically adjust its resources and management strategies, and thus maintain stable service performance in the face of demographic changes.

3. **Designing and evaluating service management strategies for improving the allocation of OR time blocks with respect to unpredictable patient arrivals.**

As a healthcare service system is complex in nature [59, 60, 61], we design and evaluate service management strategies from a self-organizing systems perspective, with the aim of improving service performance. Chapter 6 introduces this work in detail. We propose an adaptive OR time block allocation strategy incorporating historical information about OR use, focusing on ORs that provide cardiac surgery services. ORs may thus adaptively schedule their time blocks in response to any unpredictable changes in arrivals and hence achieve a trade-off between the number of bumped non-urgent surgeries and any unused urgent time blocks. To evaluate the performance of the proposed adaptive allocation strategy, we develop a multi-priority, multi-server, non-preemptive queueing model with an entrance control mechanism, to characterize the general perioperative practice of the cardiac surgery ORs in the Hamilton Health Sciences Centre in Ontario, Canada. By
applying the adaptive strategy to this queueing model, we show that our adaptive strategy is able to efficiently regulate the OR time block reservations in response to dynamically changing patient arrivals. This adaptive strategy is able to maintain a better trade-off between the number of bumped non-urgent surgeries and the number of unused urgent OR time blocks, which leads to shorter waiting lists and wait times. Furthermore, our experimental findings suggest that frequently adjusting the OR time block allocation, approximately once a month, helps to improve OR use. This finding has the potential to improve the practice of cardiac surgery services.

4. Proposing a method of behavior-based autonomy-oriented modeling for characterizing the tempo-spatial patterns in wait times, by modeling patients’ autonomous behavior and their interactions with respect to multiple impact factors.

We propose a behavior-based autonomy-oriented modeling method based on AOC [57], to understand the emergent tempo-spatial patterns in wait times. The novelty of the proposed method, compared to existing methods such as queueing theory and system dynamics, is that it models a healthcare service system from a self-organizing systems perspective. By using this method to model patients’ autonomous behavior and interactions with respect to multiple factors, we are able to uncover the essential mechanisms of positive feedback-based aggregation and collective regulation in the healthcare service system under consideration. These identified mechanisms suggest potential explanations for the dynamics and emergent patterns in patient arrivals and wait times at a systems level.

Chapter 7 presents the application of the behavior-based autonomy-oriented modeling method to uncover the underlying mechanisms for certain tempo-spatial patterns in wait times that are observed in a real-world cardiac care system. We develop an AOC-based cardiac surgery service (AOC-CSS) model to characterize the behaviors and interactions of patients and hospitals in cardiac care. We then carry out simulation-based experiments from which tempo-spatial patterns in
patient arrivals and wait times emerge. These simulated emergent patterns, especially the statistical power-law distribution of wait time variations, suggest that patient hospital selection behavior and its relationship with hospital wait times may account for the self-regulation of service utilization and wait times. The experiments also reveal that this method can be effective in explaining the self-organized regularities and investigating emergent phenomena in complex healthcare systems.

All of the findings regarding the cardiac care system described above show that our data-driven complex systems modeling approach has potential for investigating complex healthcare service systems. This approach may provide healthcare administrators and researchers with a practical way to predict the changes in service performance under different scenarios and to evaluate strategies for resource allocation and healthcare service management.

1.4 Structure of the Thesis

This thesis is organized as follows.

Chapter 2 presents background information on healthcare service systems, their complexity, and self-organization properties. It reviews existing studies and methods for empirically identifying relationships between variables and characterizing the behavior of a healthcare service system.

Chapter 3 presents the use of the SEM-based analysis method to discover the direct and moderating effects that certain geodemographic profiles exert on service utilization, focusing on cardiac surgery services in Ontario, Canada.

Chapter 4 also demonstrates the effectiveness of the SEM-based analysis method, here exploring the direct and indirect effects of a CU’s characteristics on the wait times of its subsequent SU.

Chapter 5 presents our integrated projection method for estimating the changes in service utilization and service performance with respect to demographic shifts, within the context of cardiac surgery services in Ontario.
Chapter 6 proposes and evaluates an adaptive OR time block allocation strategy that is designed from a self-organizing systems perspective to cope with unpredictable patient arrivals.

Chapter 7 presents the use of a behavior-based autonomy-oriented modeling method to find underlying working mechanisms that account for certain emergent tempo-spatial patterns in patient arrivals and wait times.

Chapter 8 summarizes this thesis and highlights possible extensions for future work.
Chapter 2

Literature Review

This chapter reviews the existing studies that are closely related to this work. We first review the basic concepts of a healthcare service system, its complexities and its self-organizational properties. This review provides the basis for us to propose a systematic approach for understanding a specific healthcare service system from a self-organizing systems perspective. We then examine related studies and the commonly used multivariate methods for understanding a healthcare service system by empirically identifying the relationships between the variables. Finally, we survey the modeling and simulation methods that have been used to characterize the behavior of specific healthcare service systems, and to unveil the underlying mechanisms that cause the emergent tempo-spatial patterns at a systems level,

2.1 Healthcare Service Systems

It is essential to view healthcare services from a complex systems perspective to understand what causes the dynamics of and relevant tempo-spatial patterns in wait times [59, 60, 61]. In this section, we review the basic concepts of a general healthcare service system, its complexity, and its self-organizational properties.
2.1.1 Basic Notions and Concepts

From a systems perspective, a healthcare service system is an open system that exchanges patients, resources, and information with the environment [62, p.32] via its input and output. Figure 2.1 presents a conceptual model, which is adapted from the open systems model [63, p.90], illustrating a healthcare service system.

![Conceptual Model](image)

**Figure 2.1:** A conceptual model illustrating a healthcare service system.

Formally, a healthcare service system $HSS$ can be characterized by a set $HSS = \{R, SS, P, ST, G\}$, where $R = \{r_1, r_2, \cdots, r_Nr\}$ corresponds to the resources within the system, including personnel and facilities, and $N_r$ is the number of dimensions representing the resources; $SS$ denotes the organizational structure of these resources; $P = \{p_1, p_2, \cdots, p_{N_p}\}$ represents the processes, referred to as the behavior of the system, and $N_p$ is the total number of processes that serve patients; $ST$ describes the states of the system at time $t$. 
\[ ST = \{st_1(t), st_2(t), \cdots, st_{N_{st}}(t)\} \], and \( N_{st} \) is the number of dimensions for measuring the states of the system; and \( G \) is the goal set of the system.

As shown in Figure 2.1, a healthcare service system can be divided hierarchically into subsystems or entities at different levels. For instance, the collection of hospitals in a region can be thought of as a healthcare service system, where each hospital can be viewed as a sub-system or a specific entity. Each sub-system can be further divided into sub-subsystems, which are units made up of the distinct personnel and facilities providing different services to patients.

All of the elements that are outside a healthcare service system and can potentially affect its input, structure, and processes are referred to as the environment \( EN \) of the system, \( EN = \{en_1(t), en_2(t), \cdots, en_{N_{en}}(t)\} \), where \( N_{en} \) is the number of elements. The system’s environment therefore includes all of the factors that affect the system and are affected by it. A healthcare service system is capable of taking in patients, resources, and information from its environment as inputs, \( IN = \{in_1(t), in_2(t), \cdots, in_{N_{in}}(t)\} \), where \( N_{in} \) is the number of different input elements. The system then processes the input in some way and returns the treated patients and information about the system states and service performance to its environment as outputs, \( OU = \{ou_1(t), ou_2(t), \cdots, ou_{N_{ou}}(t)\} \), where \( N_{ou} \) is the number of different output elements.

The output may feed back as an input into the healthcare service system and may cause changes in the system’s transformation processes and/or future outputs. The output is then referred to as a system feedback \( FE \), \( FE = \{fe_1(t), fe_2(t), \cdots, fe_{N_{fe}}(t)\} \), where \( N_{fe} \) is the number of different feedback elements. Feedback can be both positive and negative. Positive feedback can amplify the changes in the system, whereas negative feedback can regulate the changes. Feedback can affect the input and processes of the healthcare service system at different levels. For example, in response to feedback information about the service performance of each hospital, patients may make different service utilization and hospital selection decisions and thus influence the input to the
system, each hospital, and specific units. With the same feedback information, hospitals and units within the system may adjust their processes and/or reallocate their resources to improve the performance of their services in the future.

2.1.2 Complexity and Self-Organization

As recognized by some studies, a healthcare service system, as schematically illustrated in Figure 2.1, is complex and self-organizing in nature [59, 60, 61]. The complexity of a healthcare service system lies in the nature of its entities and their behavior, relationships, and interactions, [59, 60, 61, 64].

1. **Heterogeneous entities**: Entities can represent system elements, such as patients and physicians, and sub-systems, such as hospitals and units. In a healthcare service system, entities are heterogeneous due to the differences in their profiles and behavior. For instance, patients differ in terms of age, gender, ethnicity, socioeconomic background, lifestyle, decision making style, and their corresponding healthcare service utilization behavior.

2. **Autonomous and adaptive behavior**: Entities in a healthcare service system make decisions and behave rationally, based on their own behavioral rules with respect to their perceived environmental information. Their behavior can therefore adapt to a varying environment. For example, in response to regularly released wait time information about each hospital, patients may select different hospitals for specific healthcare services to avoid long wait times.

3. **Bounded rationality**: In many cases, entities can only access partial information about the system. As a result, as Herbert A. Simon stated [65], entities may have “bounded rationality” and may not exhibit optimal behavior. Patients may select a hospital that had shorter wait times in the past quarter but now has longer wait times, for example, because they do not
have access to the latest wait time information about the concerned hospitals.

4. **Coupling relationships**: Entities in a system will be constrained by structural or functional relationships, which are referred to as coupling relationships. The coupling relationships between entities can be either pre-defined or dynamically adjusting. For instance, units in a hospital have coupling relationships with one another, due to the functional constraints of the logistics of patient flows.

5. **Interactions**: Entities may spontaneously interact with multiple impact factors and other entities. During the interaction processes, entities exert changes in the environment or in other entities’ behavior. An entity may affect other entities’ behavior directly, through their coupling relationships, or indirectly, through the updated environment or through feedback.

6. **Emergence and self-organization**: Autonomous entities achieve their goals by adjusting their behavior so as to adapt to the dynamically changing environment and respond to feedback. During this self-organizing process, small changes in variables may cause larger changes in the system under some conditions. As a result, self-organized, tempo-spatial patterns may emerge from the system that are not predefined, indicating that the system is collectively regulated.

As a healthcare service system is complex and self-organizing, a growing number of studies have attempted to apply complexity science to the study of healthcare service systems. Applying complexity theory to healthcare service systems dates back to the middle of the 1990s, when Priesmeyer et al. used chaos theory, one of the classical complex systems theories [66, 67], to examine clinical pathways as nonlinear, evolving systems. Arndt and Bigelow [68] speculated on the possible applications of chaos and complexity theories for healthcare services management. Begun and White viewed the nursing profession as a complex
adaptive system and paid special attention to its inertial patterns [69]. Smethurst and Williams found that the statistical distribution of the variance in wait times to see specialists followed a power-law distribution [70], which indicated that the healthcare service system was self-organizing, possibly due to the interactions between patients and available doctors. However, to the best of our knowledge, few studies aim to understand a healthcare service system from a self-organizing systems perspective, which is the major focus of this thesis.

2.2 Empirical Identification of Relationships between Variables

Variables in a healthcare service system can have different types of relationships with one another, due to their coupling relationships and interactions. To shed light on what causes changes in wait times, studies on healthcare services have usually used multivariate analysis methods to statistically analyze empirical data, unveiling the underlying relationships between variables. In this section, we review the main kinds of relationships that healthcare services research has focused on and the multivariate analysis methods commonly used to reveal these relationships.

2.2.1 Types of Relationships

Studies on healthcare services have usually identified three types of statistical relationships between different variables.

- **Direct relationship**: A direct relationship occurs when a dependent variable (e.g., $ou_i$) is directly affected by an independent predictor variable (e.g., $in_i$ or $en_i$). The relationship is measured by the direct effect, which represents the extent to which the dependent variable changes when the predictor variable increases by one unit. Studies have found, for example, that an input variable, service utilization [26, 71], and a resource variable $r_i$ of the healthcare service...
system, the service capacity [26, 27, 72], are direct predictors of two output variables, the wait times and queue length.

- **Indirect relationship**: An indirect relationship exists when an independent variable (e.g., \(en_i\)) influences a dependent variable (e.g., \(ou_i\)) via the effect of a third variable (e.g., \(in_i\)), commonly known as a mediator variable. The indirect effect, which reflects the indirect relationship between the dependent variable (e.g., \(en_i\)) and the independent variable (e.g., \(en_i\)), is a product of the direct effect between the independent and the mediator (e.g., \(in_i\)) variables, and that between the mediator and the dependent variables. An example of an indirect relationship is between an environment variable, population size, and an output variable, wait times. A larger population can result in a greater number of patient arrivals [28, p.59], which is one of the direct antecedents of wait times [29, 72].

- **Moderating relationship**: A moderating relationship exists when the direction and/or strength of the relationship between two variables (e.g., \(en_i\) and \(in_i\)) depends on a third variable (e.g., \(en_j\)), which is known as a moderator variable. A few studies have discovered moderating effects in healthcare service systems. For instance, the prevalence of smoking and inactivity, two traditional cardiovascular risk factors, in a less-educated population [15, 22] suggests that a higher proportion of well-educated individuals in the population, one of the environment variables, may mitigate the pressure of another environment variable, population size, on an input variable, patient arrivals, and thus ease the burden on an output variable, wait times [30].

The relationship between two variables can be either linear, meaning that the changes in the dependent variable (e.g., \(ou_i\)) are proportional to the changes in the independent variable (e.g., \(en_i\) or \(in_i\)), or nonlinear, indicating that the changes in the dependent variable do not correspond with constant changes in the independent variable. Studies usually assume that the variables under
consideration are linearly related and thus use linear-model-based statistical methods, such as linear regression, principle component analysis (PCA), and factor analysis, to discover the underlying relationships. To unveil the nonlinear relationships, studies have usually used specific nonlinear functions, such as the logistic function, to transform the original data representing the dependent and independent variables, then explored the linear relationships between the transformed variables.

2.2.2 Multivariate Analysis

Three requirements must be satisfied to identify the predictors of wait times and their corresponding effects.

1. *Constructing latent and observed variables*: The variables that affect wait times can be observed or unobserved (latent constructs), as discussed in Chapter 1.2. The observed variables and LVs must therefore be modeled simultaneously.

2. *Exploring complex relationships between multiple variables*: The predictor variables for wait times may have direct, indirect, or moderating relationships with each other and with wait times, which are referred as complex relationships in this thesis. Thus, we must be able to test complex relationships between multiple variables.

3. *Supporting limited data analysis*: Publicly available data about a healthcare service system is usually aggregated and limited. We should therefore be able to explore the relationships between multiple variables using limited data.

Studies have commonly used PCA [73] to summarize a set of uncorrelated variables from empirical data [74], to extract the key predictors for wait times. PCA “converts a set of possibly correlated variables into a set of values of linearly uncorrelated variables”, called the principal components [74]. PCA is based on the assumptions that the observed variables are partially correlated; the relationships
between all of the observed variables are linear; each pair of observed variables should display a bivariate normal distribution to represent random sampling; and that the data describing the observed variables should be metric (interval/ratio) data. PCA may therefore help to transform a set of potentially correlated observed variables, e.g., \( \{en_1(t), en_2(t), \cdots, en_{N_{en}}(t)\} \), into a set of uncorrelated variables, e.g., \( \{en'_1(t), en'_2(t), \cdots, en'_{N'_{en}}(t)\} \), when analyzing a healthcare services system. PCA thus can extract potential key factors for wait times from empirical data on the environment, the input, or the healthcare service system. However, this method cannot reveal the different relationships between variables and cannot model LVs.

Unlike PCA, factor analysis aims to identify unobserved variables (i.e., LVs), called factors, that can explain the variability between a set of observed and correlated variables. Factor analysis is based on the assumptions that one or more underlying factors that can account for the variation between the given observed variables exist; variables are partially correlated; each factor is a linear construction of several observed variables; and that the data representing the observed variables should be metric data. Factor analysis has been used to extract various underlying factors in healthcare service systems, such as those that contribute to long wait times [75], and patient satisfaction with diabetes care [76] or HIV/AIDS care [77]. Factor analysis may help to extract a smaller set of LVs by removing redundancy or duplication from the correlated observed variables. However, it cannot be used to discover the different relationships between multiple LVs.

To unveil the relationships between variables from limited data, most studies have relied on multiple regression [40, 41]. Multiple regression is a general statistical method for analyzing the relationship(s) between a dependent and multiple independent variables [42]. This method consists of several types of techniques which model relationships using different linear or nonlinear equations. Studies have usually used the multiple linear regression to estimate the
contributions of different predictor variables (e.g., $en_i$, $in_i$, and $r_i$) to wait times, assuming that these variables are linearly related. For instance, researchers have used this method to explore the direct effects of hospitals’ characteristics (e.g., a university/regional hospital or a county/district county hospital) and patient socioeconomic profiles on wait times [78]. One study used multiple linear regression to examine whether old age ($\geq 65$) affects wait times in emergency departments [40]. However, this method aims to discover pairwise relationships between observed variables, and thus is not appropriate for modeling LVs or discovering indirect and moderating relationships between variables.

Some studies in healthcare have used logistic regression to deal with research questions like whether and to what extent specific variables of the environment $en_i$, input $in_i$, and/or system resources $r_i$ predict long short wait times. Logistic regression is a special type of regression that assumes that the logit of the observed dependent variable is a linear function of the observed independent variables [79]. Studies have used logistic regression to investigate, for example, whether patients less than 65 years old age and with lower education are more likely to report unacceptable wait times [41], and whether the distance from the homes of Canadian children with cancer to oncology treatment centers has a significant effect on wait times in the corresponding services [80]. Nevertheless, as the aim of logistic regression is to identify a logistic relationship between an observed dependent variable and one or more observed independent variable, this method cannot construct LVs or uncover complex relationships between variables.

In the past decade, a so-called second generation statistical technique, SEM, has drawn increasing attention. SEM enables us to investigate a series of direct, indirect, and moderating relationships between observed variables and LVs simultaneously [42]. SEM uses a measurement model and a structural model to discover the complex relationships between variables. The measurement model [42] characterizes the linear relationships between observed variables (measurement variables, MVs) and the corresponding LVs. One of the typical ways to relate MVs
to LVs is through the reflective measurement model, in which each LV is reflected in its corresponding MV. The structural model [42] describes the linear relationships between LVs. There are two classes of SEM: PLS-based SEM and covariance-based SEM [43]. PLS-based SEM is more suitable for theory building and allows for both confirmatory and exploratory modeling, while covariance-based SEM is more suitable for theory testing and is more efficient for confirmatory modeling [43]. Due to the advantages of SEM, healthcare service studies have used SEM to test a “patient satisfaction theory” in emergency departments [81]; and to investigate whether depressive symptoms are associated with glycemic control in diabetic adults and the extent to which these adults’ health behavior can explain the association [82]. SEM is applicable in this thesis, for analyzing the complex relationships between the observed variables and LVs of the environment \( EN \), input \( IN \), healthcare service system \( HSS \), and output \( OU \), such as wait times.

Based on the identified relationships between variables, healthcare administrators can predict the variations in the input or output variables if certain determinants change. The assumption for making predictions based on variable relationships is that the relationships do not change from the baseline period to the prediction period. For instance, studies have used regression models to estimate the future costs of care for cardiovascular disease from 2010 to 2030 in the United States [83], predict the mental health costs in United Kingdom [84], and estimate the medical expenditures, health care use, and mortality in Switzerland in 2010 based on the data in 2009 [85]. However, regression methods may not capture the indirect and moderating relationships between variables, which may influence the accuracy of predictions in the studies that rely on these models. These predictions also cannot demonstrate how the predicted variables change over time.

Some studies have made predictions based on time series data using the autocorrelation method. Autocorrelation describes the correlation of a random time series with itself at different time delays [86, p.459]. It makes linear predictions based on the assumption that the observed time series is self-similar.
Autocorrelations have been used, for instance, to assess the burden of children suffering from severe viral respiratory illness in an intensive care unit [87]. However, time series analysis still cannot present the dynamics of a system’s behavior.

### 2.3 Characterization of System Behavior

Studies that have used various multivariate analysis methods to explore the relationships between the variables in a healthcare service system hold the underlying assumption that the variables and entities at different levels of the system are homogeneous. Therefore, the methods used to empirically identify relationships between the variables cannot deal with the problem of how and why the output of a system changes over time. To address this problem by modeling and simulating the behavior of a healthcare service system, two requirements should be satisfied.

1. *Modeling heterogeneous and autonomous entities:* In a healthcare service system, entities (e.g., patients and hospitals) with different profiles may have distinct behavior according to their own decision making. Thus, any modeling method should consider the issue of heterogeneity and autonomy.

2. *Incorporating interactions:* In a healthcare service system, the mutual interactions between variables and entities at different levels potentially cause positive-feedback-based aggregations and nonlinear dynamics. Any modeling method should therefore incorporate the interaction issue.

To gain an understanding of the dynamics of wait times, studies have commonly used stochastic modeling and simulation, system dynamics, and ABM to model and simulate the behavior of a healthcare service system. In this section, we review these methods, their advantages, and the challenges in using them to characterize specific tempo-spatial patterns at a systems level. We present the AOC, a research
paradigm that is effective in modeling and addressing self-organization issues in complex systems, and thus may help us to uncover the mechanisms that account for the tempo-spatial patterns in wait times.

### 2.3.1 Stochastic Modeling and Simulation

In the delivery of healthcare services, patient arrivals and services to patients exhibit variability. The input of patients may randomly change over time because of unpredictable outbreaks of specific diseases and patients' autonomous behavior. Hassan et al. empirically validated the common use of the Poisson distribution to describe stochastic patient arrivals, based on recorded data on random patient arrivals in 2000 [88]. The time required for serving patients varies from one patient to another, due to the differences in patients' conditions and the severity of illnesses. The majority of studies have described stochastic services using an exponential distribution [89].

Studies have used stochastic modeling and simulation methods to model a healthcare service system by describing the stochastic input $IN$ and the processes $P$ that transform $IN$ to the output $OU$ of a healthcare service system. The aim of these methods is to estimate the probability distributions of potential outputs or states, taking into account random variation in one or more variables in the system. Queueing theory (and the corresponding queueing models) is one of the classic methods in this category for the analysis of queue lengths and waits in a system over time [50].

The origin of queueing theory may date back to the work of A.K. Erlang in the beginning of the last century [90]. Models based on queueing theory are able to mathematically analyze queue lengths and wait times in a system by specifying its random patient arrivals, random delivering services, on-site servers, and scheduling strategies. Models built using queueing theory assume that: random variables in a system statistically follow specific distributions, e.g., Poisson arrivals and exponential service rate; entities in the queues are passive, meaning that they
cannot make autonomous decisions and interact with each other; and that the system state, which is characterized by the lengths of queues, satisfies a Markov property (i.e., the future state of the system is conditional on the present state of the system, but does not rely on the past state) [91]. Based on these assumptions, this method uses a Markov chain with a transition rate matrix on a state space to describe a system.

Theoretically, we can calculate the steady-state distributions of a modeled healthcare service system with a specific queueing model and thus obtain the system’s expected queue lengths and wait times. Here, the steady-state indicates a state of equilibrium in which the distribution properties of a healthcare service system are independent of time [92]. However, in some cases, it may be difficult to use equations to describe the randomness and interdependence of certain random variables (e.g., patient arrivals and wait times), due to the coupling relationships and mutual interactions between these variables. In some complicated queueing models, it may also be difficult to mathematical analyze the steady-state distributions of the modeled system.

To address this problem, studies have used discrete event simulation, which originated around 1960 [93], to simulate queueing models. This method portrays a system’s states as a discrete sequence of events [94, 95]. An event may present a specific action (e.g., a patient joins a service waiting queue), which causes a change in the system’s state. Discrete event simulation is quite different from continuous simulation, which continuously represents the dynamics of a system over time but does not pay attention to the state changes.

To simulate a healthcare service system’s random input of patient arrivals and its processes, discrete event simulation usually integrates the Monte Carlo simulation. Discrete event simulations are often used to model deterministic systems, whereas Monte Carlo simulations sample a new value for each random variable from specific statistical distributions. Thus, a Monte Carlo simulation can effectively simulate healthcare service systems in which probability and non-determinism play a major
Stochastic modeling and simulation methods, especially queueing theory and
discrete-event simulation, lend themselves to the analysis and prediction of the
dynamic behavior of a healthcare service system. For instance, studies have used
queueing models and discrete event simulations to analyze waiting lists when
designing a specific healthcare service system [50], to present the dynamics of ORs
and recovery rooms under the constraints of capacity (e.g., beds and recovery
time) [26, 96], and to predict the performance of a healthcare service system in
different scenarios [50]. Jun et al. in 1999 [97], Fone et al. in 2003 [98], and
Jacobson et al. in 2006 [51] surveyed the application of queueing models and
discrete-event simulations in the healthcare services literature in addressing
problems such as forecasting the dynamics of patient flows with different resource
allocation strategies. Based on these reviews, in 2010, Gı́nal et al. [99] and
Cardoen et al. [100] reviewed the latest studies that use these two methods for OR
planning, scheduling, and performance modeling.

Despite the widespread application of stochastic modeling and simulation
methods in healthcare, the statistical assumptions made for the stochastic
properties are relatively strong and do not always hold in the real world. Further,
these methods assume the existence of passive entities in the system, which makes
it difficult to model entities’ autonomous behavior with respect to certain impact
factors. Therefore, these methods cannot explain how tempo-spatial patterns in
wait times emerge from individuals’ behavior and interactions.

2.3.2 System Dynamics

System dynamics is another commonly used method for modeling and simulating
healthcare service systems. It originated in the 1950s [101] and is promoted by the
System Dynamics Society\(^1\). System dynamics aims to understand the dynamically
changing behavior of a complex system by defining the interactions (which are

\(^1\)http://www.systemdynamics.org/
referred to as flows) between variables (which are referred to as stocks) that may accumulate or be exhausted over time [55]. Stochastic modeling and simulation characterize a healthcare service system by describing its stochastic properties, whereas system dynamics uses a causal loop diagram to model the internal feedback loops between the variables within a system. System dynamics assumes that: the focal system is deterministic and can be described by a set of coupled, linear or nonlinear, first-order differential or integral equations; the entities contained in a stock are homogeneous; and that the interactions between the variables, i.e., flows, are predefined and do not change during the simulation.

Studies have used system dynamics to qualitatively characterize the effects of interrelated impact factors and wait times on the cardiac care system in Ontario, Canada [102]; to model the relationships between multiple interacting diseases, healthcare service systems for delivering corresponding services, and national and state policy [103]; to identify bottle-necks in emergency healthcare by simulating patient flows [104]; and to predict the demand for ambulatory healthcare services [105]. The applications of this method in modeling various healthcare service systems are broader than those we have reviewed.

Because of its advantages in understanding the behavior of a system by modeling stocks, flows, and internal feedback loops, system dynamics provides a potentially useful means for us to investigate how the interactions between multiple variables and time delays affect the dynamics of wait times in a healthcare service system. However, system dynamics may not fully fulfill the requirements for explaining the causes of tempo-spatial patterns in wait times. The homogeneity assumption relating to stocks makes it difficult to model patients’ heterogeneous behavior, which depends on individuals’ profiles, decision making styles, and environment information. Further, the predefined, fixed interactions between stocks do not allow system dynamics to model individuals’ adaptive, autonomous behavior. Hence, this method cannot be used to investigate how tempo-spatial patterns in wait times at a systems level emerge from collective
2.3.3 Individual-Based Modeling

Studies in healthcare have also developed various system models based on individual-based modeling. Different from the stochastic modeling methods that focus on characterizing the uncertainty in a healthcare service system, and system dynamics that pays attention to feedback loops and time delays in a deterministic healthcare service system, individual-based modeling aims to describe a system by modeling and simulating the behavior of and interactions between autonomous individuals [106]. ABM, which originates from Neumann’s cellular automata machine [107] from the 1940s and Conway’s Game of Life from 1970 [108], is one of the classic individual-based modeling methods that are commonly used in healthcare services research.

ABM regards each individual as an agent, which could be either a physical element, such as a patient, or an abstract concept, such as a hospital. In ABM, each agent makes decisions individually according to its behavioral rules and perceived environmental information [109]. Agents can interact with each other through competition, cooperation, or environmental information sharing. Even a simple agent-based model can develop specific tempo-spatial patterns at a systems level, due to autonomy and interactions [110, 111]. ABM therefore enables us to explore the mechanisms that potentially explain how system behavior and certain tempo-spatial patterns arise from individuals’ behavior and the interactions between them.

Developing an agent-based model for characterizing a healthcare service system is challenging, as it requires a thorough understanding of the modeling system which is inherently complex, and there is uncertainty in designing and quantifying individuals’ behavior and interactions [112, 113]. Currently, although a unifying framework for designing, constructing, and validating agent-based models is lacking, [114, 115], several frameworks [116], or so called “meta-models” [112], have
been proposed to guide the ABM of complex systems. The frameworks are either domain-driven or pattern-orientated.

- **Domain-driven** frameworks begin with identifying and understanding the domain of the system to be modeled. Developers or modelers then build up corresponding agent-based models and conduct simulations based on the domain knowledge and a specific research context. Domain-driven frameworks, such as CoSMoS, which is proposed and promoted by the CoSMoS Project group[^117], may eliminate the uncertainties that are involved in the modeling and simulation of domain-specific systems.

- **Pattern-oriented** frameworks pay attention to discovering multiple patterns of behavior in real systems. The patterns are used to determine the modeling scope, and reduce parameter uncertainty in simulations. Pattern-oriented modeling, proposed by Grimm et al. [116, 118], is an example and is effective in modeling real systems [119].

ABM also lends itself to understanding healthcare service systems. Researchers have built different agent-based models to examine the effects of physicians’ behavior on patient outcomes [120], predict the spread of infectious diseases based on social networks [121, 122], and evaluate patient scheduling or other operation management strategies [123, 124], for example.

Although ABM provides a potentially useful means for characterizing the behavior of a system by modeling individuals’ heterogeneous behavior and interactions, it still faces several difficulties in modeling a healthcare service system to explain its emergent tempo-spatial patterns. As different agents, such as the modeled patients and hospitals, have various types of behavior and interactions in the real world, does ABM need to model the agents’ behavior and interactions as explicitly as possible? What fundamental behavior and interactions at an individual level are crucial for emerging tempo-spatial patterns at a systems level?

[^117]: [http://www.cosmos-research.org/about.html](http://www.cosmos-research.org/about.html)
and must therefore be modeled? If the model incorporates too many details, it may become too complex to assess the effects of individuals’ behavior and interactions, and other variables on the whole system. If the model omits key behavior and/or interactions, it may not capture the complex, self-organizing nature of a healthcare service system. Thus, the modeled system may not show the tempo-spatial patterns at the systems level. Few studies have successfully discovered the underlying mechanisms for the emergent patterns of a healthcare service system using ABM, which may be due to the above reasons.

The multi-agent Autonomy-Oriented Computing (AOC) [58, p.9] may be able to solve the problems facing ABM. AOC is a research paradigm that uses autonomous entities (agents) to deal with the issues of modeling and analyzing complex systems, and solving computational problems from a complex systems perspective [57, 125]. The AOC-by-prototyping technique [125] can be used to model a complex system from a self-organizing perspective. AOC-by-prototyping requires: recognizing and modeling the autonomous entities that may play significant roles in the self-organization of the system; determining and modeling the types of information that are collected and exchanged in the environment; and identifying and modeling the entities’ behavior, their direct interactions or indirect interactions via sharing information in the environment, and the positive or negative feedback loops, which may enable the system to exhibit collective aggregations or regulations. AOC-by-prototyping should be a recursively trial-and-error process to make the synthetic system as vivid as possible. During this process, some parameters are initialized and configured to make the synthetic model approximate the real system more closely. The final synthetic model can be used to reveal the underlying mechanisms of positive-feedback-based aggregations or negative-feedback-based regulations, which may account for the observed self-organization and emergent behavior of the real system.

Due to the advantages of modeling a system from a self-organizing perspective, AOC’s effectiveness has been validated in a variety of real-world applications, such as
understanding the dynamics of the interactions between HIV and the human immune system [126]. AOC therefore offers a method for developing a specific healthcare service model to characterize the dynamics of and tempo-spatial patterns in wait times.

2.4 Summary

This chapter reviewed the literature on (1) the basic ideas of a healthcare service system, its complexity and its self-organizing properties; (2) the types of relationships between variables in the system and the methods that can be used to empirically identify these relationships from limited data; and (3) the modeling and simulation methods for characterizing the behavior of a system. Based on this review, we present the motivations for developing the data-driven complex systems modeling approach that consists of an SEM-based analysis, integrated projection, designing and evaluating service management strategies, and behavior-based autonomy-oriented modeling to understand the nature of a complex healthcare service system in terms of wait times. We can also investigate the differences between the methods used in existing studies and those considered in this thesis.
Chapter 3

Discovering the Effects of Demand Factors: On Service Utilization of a Single Unit

This chapter presents the application of an SEM-based analysis method to explore the complex effects of certain geodemographic profiles (as specific demand factors) on service utilization. Figure 3.1 illustrates how the research focus of this chapter fits into the context of understanding a healthcare service system. We explore whether population size, age profile, service accessibility, and educational profile have direct or moderating effects on service utilization in cardiac care services in Ontario, Canada. We propose our research hypotheses and a conceptual model based on a thorough literature review. We test these hypotheses using SEM based on aggregated data representing Ontario’s geodemographic profiles and patient arrivals for cardiac surgery services between 2004 and 2007, obtained from the CCN. We discuss and interpret our findings and make practical suggestions for improving wait time management.
Figure 3.1: A schematic diagram illustrating the use of an SEM-based analysis to explore the complex effects of certain geodemographic profiles on healthcare service utilization. (a) The research focus of this chapter (highlighted in red) is part of the larger context of understanding a healthcare service system. (b) Exploring underlying effects using an SEM-based analysis.

### 3.1 Introduction

Geodemographic factors, such as population size [28], age [14, 127], geographic accessibility to services [20], and level of education [128, 129], have been recognized as important determinants of healthcare service utilization [36, 37]. Geodemographic factors are conventionally used to estimate healthcare needs (e.g., the population needs-based funding formula [130]) to develop better resource allocation and shorter wait times. The majority of studies have focused on examining pair-wise relationships between geodemographic factors and healthcare service utilization, with a scarcity of research exploring how the geodemographic
factors interact to affect healthcare service utilization.

Nevertheless, as previous studies have suggested [15, 20, 131], certain geodemographic factors may moderate (i.e., change the direction and/or strength of) [42] the effects that other geodemographic factors have on healthcare service utilization. For instance, if one area has more healthcare service providers, the burden of population growth and aging on the patient arrivals for a specific hospital in that area may be alleviated, as patients residing there have more choices and thus will be more likely to be distributed among multiple hospitals. Geographic accessibility to services (referred to hereafter as service accessibility) [20] may therefore have moderating effects on the relationships between a population’s size, age profile, and service utilization. As an additional example, individuals, including seniors, with different education backgrounds may have different lifestyles [15] that can influence their risk for cardiovascular disease [128, 129] and their healthcare service utilization behavior [131]. The educational profile may therefore have a moderating effect on the relationship between population size and healthcare service utilization.

To the best of our knowledge, no previous studies have explored the potential effects of geodemographic factors, such as service accessibility and education level, in moderating the influence of other geodemographic factors, such as population size and age profile, on healthcare service utilization. To fill this gap in the literature, we aim to examine both the direct and moderating effects of geodemographic profiles on service utilization within the context of cardiac care, in various sub-regions of Ontario, Canada. The sub-regions of concern are local health integration networks (LHINs)¹, which differ from one another in terms of their administrative areas, geographic sizes, and geodemographic profiles (as shown in Table 3.1). Although LHINs have been in operation for years, there is a lack of academic research examining how their geodemographic profiles affect healthcare service utilization.

¹http://www.lhins.on.ca/home.aspx
To achieve our objective of examining the direct and moderating effects of geodemographic profiles, we develop hypotheses based on the literature review and construct a corresponding conceptual model. We then test the model using SEM [42, 43], based on publicly available aggregated data representing the geodemographic factors and cardiac surgery service utilization in Ontario from 2004 to 2007.

3.2 The Effects of Geodemographic Profiles

In this study, we explore how geodemographic factors interact to influence healthcare service utilization in the context of cardiac surgery services. According to the literature, the commonly considered geodemographic factors include the population size, age profile, service accessibility, educational profile. In this section, we review the literature and develop hypotheses regarding the effects of geodemographic profiles (as direct antecedents and moderators) on healthcare service utilization.

3.2.1 Hypotheses

(1) The Direct Effect of Population Size on Service Utilization

Population size, representing the total population that may use the cardiac surgery services in an LHIN, has been shown to exert a direct positive influence on service utilization, which is represented by the number of patient arrivals. A larger population may translate into a greater number of people using healthcare services to prevent or treat various types of illnesses [28, p.59]. Population growth, which can produce more cardiovascular patients, has been identified as one of the major driving forces behind changes in the number of patient arrivals [102]. We thus hypothesize that:

Hypothesis 1 (H1): Population size has a direct positive effect on service utilization.
Table 3.1: The names, sizes, and scopes of LHINs in Ontario, Canada

<table>
<thead>
<tr>
<th>ID</th>
<th>LHIN name</th>
<th>Area ($km^2$)</th>
<th>PD</th>
<th>Boundary (Major cities/towns/counties)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Erie St. Clair</td>
<td>7323.7</td>
<td>86.1</td>
<td>Windsor, Lambton, Chatham-Kent, and Essex</td>
</tr>
<tr>
<td>2</td>
<td>South West</td>
<td>20903.5</td>
<td>43.1</td>
<td>London, Stratford, Elgin, Middlesex, Oxford, Perth, Huron, Bruce, and part of Grey</td>
</tr>
<tr>
<td>3</td>
<td>Waterloo Wellington</td>
<td>4746.6</td>
<td>144.6</td>
<td>Wellington, Waterloo, Guelph, and part of Grey</td>
</tr>
<tr>
<td>4</td>
<td>Hamilton Niagara Haldimand Brant</td>
<td>6473.0</td>
<td>203.3</td>
<td>Hamilton, Niagara, Haldimand, Brant, and parts of Halton and Norfolk</td>
</tr>
<tr>
<td>5</td>
<td>Central West</td>
<td>2590.0</td>
<td>285.7</td>
<td>Dufferin, parts of Peel, York, and Toronto</td>
</tr>
<tr>
<td>6</td>
<td>Mississauga Halton</td>
<td>1053.7</td>
<td>956.7</td>
<td>Mississauga, parts of Toronto, Peel, and Halton</td>
</tr>
<tr>
<td>7</td>
<td>Toronto Central</td>
<td>192.0</td>
<td>5678.9</td>
<td>A large part of Toronto</td>
</tr>
<tr>
<td>8</td>
<td>Central</td>
<td>2730.5</td>
<td>561.3</td>
<td>Parts of Toronto, York, and Simcoe</td>
</tr>
<tr>
<td>9</td>
<td>Central East</td>
<td>15274.1</td>
<td>93.8</td>
<td>Durham, Kawartha Lakes, Haliburton, Highlands, Heterborough, parts of Northumberland, and Toronto</td>
</tr>
<tr>
<td>10</td>
<td>South East</td>
<td>17887.2</td>
<td>26.1</td>
<td>Kingston, Hastings, Lennox and Addington, Prince Edward, and Frontenac</td>
</tr>
<tr>
<td>11</td>
<td>Champlain</td>
<td>1763.1</td>
<td>65.1</td>
<td>Ottawa, Renfrew, Prescott and Russell, Stormont, and Dundas and Glengarry</td>
</tr>
<tr>
<td>12</td>
<td>North Simcoe Muskoka</td>
<td>8372.3</td>
<td>50.5</td>
<td>Muskoka, parts of Simcoe and Grey</td>
</tr>
<tr>
<td>13</td>
<td>North East</td>
<td>395576.7</td>
<td>1.4</td>
<td>Nipissing, Parry Sound, Sudbury, Algoma, Cochrane, and part of Kenora</td>
</tr>
<tr>
<td>14</td>
<td>North West</td>
<td>406819.6</td>
<td>0.6</td>
<td>Thunder Bay, Rainy River, and most of Kenora</td>
</tr>
</tbody>
</table>

(2) The Direct Effect of Age Profile on Service Utilization

Age profile, here defined as the proportion of seniors (i.e., those older than 50) in the population that may use the cardiac surgery services in an LHIN, has been recognized as another important factor that may influence service utilization. Old age is a traditional cardiovascular risk factor [132]. Other risk factors for cardiovascular disease, such as hypertension, obesity, and physical inactivity, are also more prevalent in the segment of the population aged 50 and above [133, 134]. Further, age groups vary in their healthcare service utilization behavior [14, 127], with seniors typically exhibiting a higher rate of use. A larger senior population may therefore result in more cardiovascular patients [135], leading to a greater number of patient arrivals for healthcare services, such as cardiac surgery [102]. Therefore, we hypothesize that:

Hypothesis 2 (H2): Age profile has a direct positive effect on service utilization.

(3) The Moderating Effects of Service Accessibility

Geographic accessibility to healthcare services in an area (i.e., service accessibility) is an important factor influencing patients’ decisions regarding the use of such services [19, 20, 34]. Seidel et al. [19] found that patients’ willingness to use healthcare services was negatively associated with the distance between their residences and the destination hospital. A survey conducted by the CCN [34] showed that the driving distance between home and a hospital was one of the most important factors for patients in choosing a specific hospital, and that more than 80% of cardiovascular patients were not willing to visit hospitals far from home. Extending these findings, we conjecture that if there are several accessible hospitals in one area, patient arrivals for any one particular hospital may decrease, as the difference in the time needed for patients to travel to one hospital or another is negligible. Under such circumstances, we would expect patients to be dispersed among several hospitals, resulting in reduced wait times at any particular hospital in the area.
In this study, higher service accessibility for an LHIN implies that residents in that LHIN have access to more possible healthcare service providers. As a result, the number of patient arrivals at any one hospital in the LHIN may decrease. The pressure of population size or the age profile on each of the hospitals in an LHIN with higher service accessibility may be mitigated, because patients (including seniors) in that LHIN are likely to be dispersed among several hospitals. Thus we hypothesize that:

Hypothesis 3.1 (H3.1): Service accessibility has a direct negative effect on service utilization.

Hypothesis 3.2 (H3.2): Service accessibility has a negative moderating effect on the relationship between the population size and service utilization.

Hypothesis 3.3 (H3.3): Service accessibility has a negative moderating effect on the relationship between the age profile and service utilization.

(4) The Moderating Effects of Educational Profile

Educational profile is defined as the proportion of well-educated individuals (i.e., those with more than a high school education) in the population that may use the cardiac surgery services in an LHIN and is an important factor that may also affect healthcare service utilization. Individuals with different education backgrounds manifest different lifestyles [15] and are thus associated with different levels of risk for cardiovascular disease [128, 129] and different service utilization behavior [131]. For instance, a longitudinal secondary data study in Canada showed that smoking and inactivity, two traditional cardiovascular risk factors, were more prevalent in the less well-educated (senior) population [15]. This study suggested that people in the less well-educated group might have a higher demand for healthcare services related to cardiovascular disease. Another study showed that diabetic patients who were at greater risk for cardiovascular disease were more willing to perform self-care behavior if they were well-educated [131]. These findings suggest that, in addition to directly affecting service utilization, a higher
proportion of well-educated individuals in the population may mitigate the pressure of population size and aging on service utilization. Thus, we hypothesize that:

Hypothesis 4.1 (H4.1): *Educational profile* has a direct negative effect on *service utilization*.

Hypothesis 4.2 (H4.2): *Educational profile* has a negative moderating effect on the relationship between *population size* and *service utilization*.

Hypothesis 4.3 (H4.3): *Educational profile* has a negative moderating effect on the relationship between *age profile* and *service utilization*.

### 3.2.2 The Conceptual Model

The research model, presented in Figure 3.2, illustrates the hypothesized relationships to be tested in this study.

### 3.3 SEM Tests and Results

#### 3.3.1 Geodemographic and Service Administrative Data in Ontario

We use aggregated data from 2004 to 2007, obtained from Statistics CCN, to test the hypothesized relationships. Geodemographic data with respect to *population size*, *age profile*, and *educational profile* are gathered from Statistics Canada. According to the census data released by Statistics Canada, the geodemographic changes from year to year in each LHIN are rather gradual. For instance, between the 2001 and 2006 censuses, the population in Ontario grew by approximately 6.6% [33]. Thus, it is reasonable to assume that the 2006 Canadian census\(^2\) will approximately reflect the geodemographics of Ontario between 2004 and 2007. Based on the 2006 Canadian census data\(^3\), we select 47 major cities and towns in

\(^2\)http://www12.statcan.ca/census-recensement/index-eng.cfm  
\(^3\)http://www.statcan.gc.ca/estat/estat-eng.htm
Ontario with populations of more than 40,000 to derive the geodemographic profiles for 14 LHINs. The 40,000 population cut-off point ensures that the cities and towns included in our study represent approximately 90.72% of Ontario’s population (shown in Figure 3.3). Patients residing in an LHIN may travel to other LHINs to receive cardiac surgeries. For example, 25% of patients residing in the Central West LHIN received treatment from hospitals in the Missisauga Halton LHIN in the 2007/2008 fiscal year [136]. We therefore estimate the population that could potentially use the cardiac surgery services in each LHIN, including those residents living in other LHINs, and thereafter derive the corresponding geodemographic profiles.
Figure 3.3: Population distribution across the selected cities and towns in Ontario. The populations of cities and towns in Ontario follow a power-law distribution (correlation coefficient $R = -0.922$, standard deviation $SD = 0.2441$, $p < 0.0001$). The selected cities and towns have populations larger than 40,000 and cover approximately 90.72% of Ontario’s population.

The measurement value for \( \text{population size} \) of LHIN \( i \) is calculated by:

\[
PS'_i = \sum_{j=1}^{14} PS_j PD_{ji} \quad (i, j \in [1, 14], i \neq j),
\]

where \( PS'_i \) denotes the measurement value of \( \text{population size} \) in LHIN \( i \), \( PS_j \) represents the population size in LHIN \( j \), and \( PD_{ji} \) is the proportion of patients residing in LHIN \( j \) but receiving services in LHIN \( i \). The data representing \( PD_{ji} \) were obtained from [136].

The measurement values for \( \text{age profile} \) and \( \text{educational profile} \) for LHIN \( i \) are calculated by:

\[
V'_i = \frac{\sum_{j=1}^{14} V_j PD_{ji}}{PS'_i} \quad (i, j \in [1, 14], i \neq j),
\]
where $V_i'$ denotes either the proportion of the senior/well-educated population in LHIN $i$; $V_j$ denotes the number of people aged 50 and above, or the number of well-educated people in LHIN $j$, respectively; $PD_{ji}$ is the proportion of patients residing in LHIN $j$ but receiving services in LHIN $i$; and $PS_i'$ is the measurement value of population size in LHIN $i$.

We operationalize service accessibility as the proportion of the population residing within a 30-minute driving time to the nearest hospitals providing cardiac surgery services in an LHIN [137]. The 30-minute driving time is selected as a threshold to measure healthcare service accessibility in accordance with previous work [138, 139] and the CCN’s recommendations [140]. The driving time from each selected city or town to the nearest hospital that provides cardiac surgery services is estimated using the “Get directions” function in Google Maps\(^4\). In Google Maps, a city or town is represented by the central point of its polygonal area\(^5\). Unlike a geographical information system (GIS), which estimates driving time based on the lengths of roads and road speed limits [141, 142], Google Maps considers the actual traffic conditions on roads. Hence, Google Maps may provide a more realistic driving time than a GIS. As there may be several routes between a city or town and a hospital in Google Maps, we tabulate the driving time for each selected city or town to all of the hospitals providing cardiac surgery services and select the route with the shortest driving time to approximate the service accessibility for the LHINs. Service accessibility is calculated by:

$$SA_i = \frac{\sum_{k=1}^{K_i} PS_{ki} \ast \delta_{ki}}{PS_i}, \quad (3.3)$$

where $SA_i$ is the service accessibility of LHIN $i$; $PS_{ki}$ is the population size of a city/town $k$ in LHIN $i$; $K_i$ is the number of cities/towns selected in LHIN $i$; $PS_i$ is the population size of LHIN $i$; and $\delta_{ki}$ is a parameter denoting whether a city/town $k$ in LHIN $i$ is within a 30-minute driving time to the nearest hospital. If the driving time from a city/town $k$ in LHIN $i$ to its nearest hospital is within 30 minutes,

\(^4\)https://maps.google.com

\(^5\)https://developers.google.com/maps/documentation/staticmaps/
\( \delta_{ki} = 1; \) otherwise, \( \delta_{ki} = 0. \)

The geodemographic profiles for the various LHINs are summarized in Table 3.2.

Table 3.2: The measurement values for the geodemographic profiles of LHINs providing cardiac surgery services (2006)

<table>
<thead>
<tr>
<th>ID</th>
<th>LHIN name</th>
<th>( PS'_i )</th>
<th>( A'_i ) (%)</th>
<th>( SA_i ) (%)</th>
<th>( E'_i ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>South West</td>
<td>762804</td>
<td>32.55</td>
<td>41.05</td>
<td>62.68</td>
</tr>
<tr>
<td>3</td>
<td>Waterloo Wellington</td>
<td>671709</td>
<td>29.73</td>
<td>77.69</td>
<td>64.16</td>
</tr>
<tr>
<td>4</td>
<td>Hamilton Niagara Haldimand</td>
<td>796559</td>
<td>33.83</td>
<td>51.54</td>
<td>61.25</td>
</tr>
<tr>
<td></td>
<td>Brant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Mississauga Halton</td>
<td>912292</td>
<td>27.54</td>
<td>88.20</td>
<td>71.51</td>
</tr>
<tr>
<td>7</td>
<td>Toronto Central</td>
<td>3813418</td>
<td>29.97</td>
<td>100.00</td>
<td>70.12</td>
</tr>
<tr>
<td>8</td>
<td>Central</td>
<td>637510</td>
<td>30.07</td>
<td>75.13</td>
<td>69.35</td>
</tr>
<tr>
<td>10</td>
<td>South East</td>
<td>198366</td>
<td>33.90</td>
<td>65.10</td>
<td>66.37</td>
</tr>
<tr>
<td>11</td>
<td>Champlain</td>
<td>651966</td>
<td>32.80</td>
<td>86.40</td>
<td>74.16</td>
</tr>
<tr>
<td>13</td>
<td>North East</td>
<td>189353</td>
<td>37.32</td>
<td>37.27</td>
<td>61.37</td>
</tr>
</tbody>
</table>

\( PS'_i \): the measurement value for population size of LHIN \( i \); \( A'_i \): the measurement value for age profile of LHIN \( i \); \( SA_i \): the measurement value for service accessibility of LHIN \( i \); \( E'_i \): the measurement value for educational profile of LHIN \( i \).

The data representing cardiac surgery service utilization from 2004 to 2007 is obtained from the CCN\(^6\). The CCN is a provincial system that includes 11 hospitals providing cardiac surgery services in Ontario and provides quarterly statistical data on the waiting queue length and the number of completed surgery cases in a month at each hospital. Based on the CCN data, the average number of cardiac surgery patient arrivals in a hospital \( i \) each month over a quarter \( t \) (\( Arrival'_i \)) can be calculated by adding the number of completed cases to the number of patients waiting in the queue (\( NoWait'_i \)), and subtracting the waiting

\(^6\)http://www.ccn.on.ca/
queue length at time $t-1$ ($\text{NoWait}_{i}^{t-1}$). An overview of the aggregated data on service utilization for each hospital is shown in Table 3.3.

Table 3.3: Cardiac surgery service utilization from 2004 to 2007 in Ontario hospitals

<table>
<thead>
<tr>
<th>LHIN ID</th>
<th>Hospital</th>
<th>Service utilization (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>London Health Sciences Centre</td>
<td>111</td>
</tr>
<tr>
<td>3</td>
<td>St. Mary’s General Hospital</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>Hamilton Health Sciences</td>
<td>112</td>
</tr>
<tr>
<td>6</td>
<td>Trillium Health Centre</td>
<td>86</td>
</tr>
<tr>
<td>7</td>
<td>St. Michael’s Hospital</td>
<td>88</td>
</tr>
<tr>
<td>7</td>
<td>Sunnybrook Health Sciences Centre</td>
<td>71</td>
</tr>
<tr>
<td>7</td>
<td>University Health Network</td>
<td>143</td>
</tr>
<tr>
<td>8</td>
<td>Southlake Regional Health Centre</td>
<td>64</td>
</tr>
<tr>
<td>10</td>
<td>Kingston General Hospital</td>
<td>53</td>
</tr>
<tr>
<td>11</td>
<td>University of Ottawa Heart Institute</td>
<td>91</td>
</tr>
<tr>
<td>13</td>
<td>Hôpital Régional de Sudbury</td>
<td>38</td>
</tr>
</tbody>
</table>

Note: Service utilization is measured by the number of patient arrivals at a hospital in a month.

### 3.3.2 Two-Step SEM Tests

The partial least squares (PLS)-based SEM software SmartPLS\(^7\) is used to test the hypothesized relationships. PLS-based SEM, when compared with LISREL, another major type of SEM, has the advantage of theory development and thus is more appropriate for exploratory modeling [43]. In this study, all of the LVs (population size, age profile, service accessibility, educational profile, and service utilization) are

\(^7\)http://www.smartpls.de/
modeled as reflective constructs, which are constructs viewed as causing, as opposed to being caused by, the observed variables [143].

We conduct a two-step test to test both the hypothesized direct and moderating effects.

- Step 1: Test the direct effects of population size and age profile on healthcare service utilization;
- Step 2: Explore the direct and the moderating effects of educational profile and service accessibility on service utilization.

### 3.3.3 Test Results

The research hypotheses are tested using secondary data on the service utilization of cardiac surgery in Ontario and the relevant geodemographic factors between 2004 and 2007 (16 quarters). The mean and standard deviation of the variables are summarized in Table 3.4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>784907</td>
<td>367484</td>
<td>189353</td>
<td>1271139</td>
</tr>
<tr>
<td>Age profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50+%</td>
<td>31.60</td>
<td>2.63</td>
<td>27.54</td>
<td>37.32</td>
</tr>
<tr>
<td>Service accessibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤30' %</td>
<td>67.90</td>
<td>19.59</td>
<td>37.27</td>
<td>100.00</td>
</tr>
<tr>
<td>Educational profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;High school %</td>
<td>67.38</td>
<td>4.24</td>
<td>61.25</td>
<td>74.16</td>
</tr>
<tr>
<td>Service utilization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. patient arrivals in a month</td>
<td>82</td>
<td>34</td>
<td>16</td>
<td>211</td>
</tr>
</tbody>
</table>

Table 3.4: Summary statistics for the geodemographic factors and cardiac service utilization in Ontario between 2004 and 2007
(1) Measurement Model

The common evaluation metrics for model fitting in PLS-based SEM are Cronbach’s alpha, construct reliability, and average variance extracted. As we use one observed variable for each LV, both Cronbach’s alpha and the construct reliability of each LV are equal to 1, suggesting that all of the LVs are internally consistent [42]. The average variance extracted for each LV is also equal to 1, indicating adequate convergent validity [42]. Moreover, the correlations between each LV and the other LVs are smaller than the square root of the average variance extracted, indicating adequate discriminant validity [144].

(2) The Effects of Population Size and Age Profile on Service Utilization

As Figure 3.4 reveals, in support of H1 and H2, both population size and age profile have significant positive effects on service utilization, with path coefficients of $\beta = 0.737$ ($t = 13.205, p < 0.01$) and $\beta = 0.284$ ($t = 5.051, p < 0.01$), respectively. These results support the previous findings that a larger population size [28, 102] and a greater proportion of residents older than 50 [133, 134] in a geographic area imply more cardiac surgery patients in the hospital(s) of that area.

![Figure 3.4: SEM test results: the effects of population size and age profile on service utilization.](image-url)
(3) The Effects of Service Accessibility and Educational Profile on Service Utilization

As Figure 3.5 shows, in support of H3.1 and H3.2, service accessibility is negatively related to service utilization ($\beta = -0.210, t = 2.101, p < 0.01$), and weakens the effect of population size on service utilization ($\beta = -0.606, t = 5.240, p < 0.01$). The findings suggest that the more accessible an LHIN is in terms of healthcare services (i.e., the more individuals within a 30-minute driving time to the nearest hospital providing cardiac surgery services), the fewer the patient arrivals at any one hospital in this LHIN and the weaker the effect of population size on service utilization. However, H3.3 is not supported by our data ($\beta = -0.070, t = 0.661, p > 0.05$), indicating that service accessibility does not have a moderating effect on the relationship between age profile and service utilization.

![Figure 3.5: SEM test results: service accessibility as a moderator.](image)

H4.1 is not supported by our data ($\beta = 0.050, t = 1.088, p > 0.1$), as shown in Figure 3.6, suggesting that educational profile does not have a direct effect on patient service utilization for cardiac surgery. However, in support of H4.2 and
H4.3, our results reveal that educational profile weakens the effects of population size and age profile on service utilization, with path coefficients of $\beta = -0.595$ ($t = 7.592, p < 0.01$) and $\beta = -0.286$ ($t = 4.987, p < 0.01$), respectively. The effects of population size and age profile on service utilization in a well-educated LHIN is therefore probably not as strong as in a less well-educated LHIN.

Table 3.5 summarizes the testing results for each of the hypotheses.

Table 3.5: Hypothesis testing results

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Supported?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1, H2, H3.1, H3.2, H4.2, H4.3</td>
<td>Fully supported</td>
</tr>
<tr>
<td>H3.3, H4.1</td>
<td>Not supported</td>
</tr>
</tbody>
</table>

**3.4 Discussion**

Meeting the needs of a population is one of the most important considerations when allocating healthcare resources in Canada, and worldwide [130]. Previous
research has advocated the allocation of resources according to the needs of the population, as assessed by an estimation method [130] that considers demographic-based indicators (e.g., age, education, and smoking) [145, 146]. However, Kephart and Asada [146] noted substantial differences between estimated and real service needs in some regions when examining traditional estimation methods. The needs estimation method may simply be a linear combination of all of the considered factors that does not consider how these factors interact with one another, resulting in a biased estimation. Therefore, an in-depth understanding of the direct and moderating interactions between the geodemographic factors and healthcare service utilization may suggest better estimation methods for healthcare service needs. LHINs are sub-provincial administrative units responsible for planning and funding healthcare services for their corresponding geographic areas in Ontario. By uncovering interesting relationships between LHINs’ geodemographic factors and healthcare service utilization, we provide LHIN administrators with valuable information to consider in their planning and/or managing of healthcare service resources.

In this chapter, we demonstrated that service accessibility has a significant moderating effect on the population size-service utilization relationship, and that educational profile exerts significant moderating effects on both the population size-service utilization relationship and the age profile-service utilization relationship. These relationships have not been reported previously. The results of our analysis confirm our prediction that service accessibility is negatively associated with service utilization, and that it weakens the effect of population size on service utilization. The results suggest that the more healthcare services are accessible in an area, the fewer cardiac surgery patient arrivals any one hospital in that area will have. We consider the Hamilton Niagara Haldimand Brant LHIN (LHIN 4) and its neighbor, the Mississauga Halton LHIN (LHIN 6), as examples. In 2007, the proportion of patients receiving cardiac surgery services in their

---

8 http://www.lhins.on.ca/home.aspx
resident LHINs (referred to as the inside-LHIN proportion) was 82% in LHIN 4 and 72% in LHIN 6 [136], whereas the service accessibility was approximately 51.54% in LHIN 4 and 88.20% in LHIN 6, as shown in Table 3.2. As both LHIN 4 and LHIN 6 have only one hospital in their own areas, the higher accessibility of LHIN 6 compared to LHIN 4 suggests that there are more accessible hospitals in the LHINs surrounding LHIN 6 than in those surrounding LHIN 4. As a result, patients dwelling in LHIN 4 are less likely to visit hospitals in other LHINs, compared to those dwelling in LHIN 6, and thus the inside-LHIN proportion for LHIN 4 is higher than that for LHIN 6. Accordingly, we expect that for LHINs with better accessibility to cardiac surgery services (e.g., LHINs 3, 6, 7, and 11, as shown in Table 3.2), the pressure of population growth in each of these LHINs on the hospital(s) within the LHIN may decrease.

In contrast, the negative but insignificant moderating effect of service accessibility on the relationship between the age profile and service utilization may be because older people are more willing to visit a familiar hospital or a hospital with familiar physicians [34]. Consequently, service accessibility in an LHIN, which reflects patients’ options in healthcare services, may have little effect on the senior population’s decisions when choosing cardiac surgery services.

The negative moderating effects of educational profile suggest that the effect of population size and age profile on service utilization is less pronounced in a well-educated population than it is in a less-educated population. Well-educated individuals, including the elderly, may have healthier lifestyles [15] and are more inclined to receive routine physical examinations and engage in self-care behavior [131]. Consequently, they are less likely to develop severe cardiovascular disease that requires cardiac surgery services [22]. As illustrated in Table 3.4, the educational profiles of the LHINs in 2006 vary from 61.25% to 74.16%, with a mean value of 67.38% and a standard deviation of 4.24%. The effects of population growth and aging on patient arrivals in each LHIN may therefore vary depending on the educational profiles of that LHIN. As shown in Table 3.2, LHINs 6, 7, 8,
and 11, which have more educated populations (indicated by higher-than-average educational profiles), may have lower patient arrivals due to population growth and aging, compared to other LHINs.

Previous research has identified population growth and aging as two important factors driving the need for healthcare services in Ontario [23], and thus affecting patient arrivals. Likewise, our findings reveal a significant relationship between population size and service utilization, and between age profile and service utilization. This finding suggests that, monitoring the trends in population growth and aging is an effective precautionary approach for healthcare administrators aiming to provide sustainable healthcare services.

The literature has noted the significant positive effect that service utilization exerts on the important performance indicator of hospital wait times [29, 26]. Our findings suggest that geodemographic factors, such as population size, age profile, service accessibility, and educational profile, may indirectly affect wait times for cardiac surgery services via their influence on patient arrivals. Therefore, healthcare administrators should consider the roles of the geodemographic factors in their efforts to improve wait times for healthcare services.

This study concentrates on four specific geodemographic factors, i.e., population size, age profile, service accessibility, and educational profile. These factors are identified based on a literature review and are not significantly correlated as tested on our aggregated data. It should be noted that there may be other geodemographic factors influencing service utilization, such as income, one of the commonly considered geodemographic characteristics in the literature. In this work, we do not pay attention to those factors because (1) some of them may significantly correlate with the population size, age profile, service accessibility, or educational profile (e.g., education attainment and income in a population have a causal relationship, as indicated by a few studies [147]); (2) the SEM test results show that approximately 50% of the variability in service utilization can be explained, suggesting that only considering the four factors in the data test is
acceptable for they account for the major part of the variance in service utilization.

### 3.5 Summary

In this chapter, we demonstrated how to use the SEM-based analysis to explore the moderating effects of certain geodemographic factors on healthcare service utilization, in addition to examining the direct effects of these geodemographic factors. Unlike previous research, we used an SEM technique and aggregated data on geodemographic factors and cardiac surgery services in Ontario, Canada to test the hypothesized relationships. The results reveal that geodemographic changes due to population growth and aging may significantly affect cardiac surgery service utilization. Geographic accessibility to healthcare services and a population’s educational profile exert significant effects on patient arrivals for cardiac surgery services, both as direct antecedents and as moderators. Our findings suggest the importance of considering the geodemographic profiles of a geographic area, and sometimes its neighboring areas, when allocating healthcare service resources, to strategically improve service utilization and reduce wait times. Finally, the work presented in this chapter demonstrates that the SEM-based analysis can be used to empirically identify the complex relationships between certain demand factors and wait times.
Chapter 4

Discovering the Effects of Supply Factors: On Wait Times across Units

This chapter presents the use of an SEM-based analysis to explore whether and how the characteristics of one unit (such as the supply factors service utilization, capacity, throughput, and wait times) affect the wait times of subsequent units in a hospital. Figure 4.1 illustrates how the research focus of this chapter fits into the context of understanding a healthcare service system. We explore the wait time relationship between units using the CU and the SU, two closely related units that are networked via patient flow, in cardiac care services in Ontario, Canada. We propose research hypotheses and a corresponding conceptual model based on a literature review. We test these hypotheses using SEM based on aggregated data representing the characteristics of the two units in 11 hospitals in Ontario between 2005 and 2008. We finally discuss the interpretation of and possible extensions to our findings.
Figure 4.1: A schematic diagram illustrating the use of an SEM-based analysis to examine the effect of a preceding unit’s characteristics on the wait times of a subsequent unit. (a) The research focus of this chapter (highlighted in red color) is part of the larger context of understanding a healthcare service system. (b) Exploring underlying effects using an SEM-based analysis.

4.1 Introduction

The effect of highly fluctuating service utilization (represented by the number of patient arrivals) and available service capacity on the performance of a healthcare service system has been long deserving of attention [148, 149]. Service utilization, capacity, and performance are all important characteristics describing a healthcare service system. Service utilization is often represented by the number of visits to services [150, 151] or the expenditure on services [152, 153]. Some of the factors affecting the service utilization of a healthcare service system are increasing numbers of patients due to aging and a growing population [102], the incidence of specific diseases such as diabetes [154], the development of diagnostic and
treatment technology [102], patient status such as the seriousness of the illness [155], the position of the patient on a waiting list [156], the geographic distance between the patient and the services [19], patients’ personal profiles (e.g., demographics [157] and socioeconomic condition [16, 158]), and unpredictable patient behavior like balking, reneging, jockeying, and repeating [96, 97, 159].

The capacity of a healthcare service system denotes the resources (e.g., financial, human, and physical) available to serve patients [160]. Capacity is usually judged by the quantity and quality of the resources at hand [27, 102] or the working time available [161]. Capacity is affected by factors such as human resources, for example skilled doctors and assistants (e.g., nurses, anesthetists) [27]; physical resources, for example beds and equipment [102]; management strategies, for example resource utilization and allocation [162]; and resource planning and scheduling [97, 162].

Performance is commonly summarized using throughput and wait times [24, 97, 163]. Throughput is typically quantified by counting the number of patients who have received a needed healthcare service in a given period [164]. It is thus a way to observe the use of health care service resources. Unlike throughput, wait times represent the amount of time patients have to wait before receiving needed healthcare services [24, 165]. Wait times are a particular concern in healthcare, especially for key services such as catheterization and cardiac surgery. A long wait is not only an impediment to quality care but also a risk factor for patients [3, 166]. There are various measurements for wait times, such as median wait time (i.e., the point at which half of the patients have received their treatment and the other half are still waiting) and queue length (i.e., the total number of patients in the waiting list) [24, 165]. Wait times differ depending on patient urgency categories. In a government dominated healthcare service system (e.g., Canada), each patient on the key units’ waiting lists is assigned an urgency rating score according to the presenting symptoms [167, 168]. Wait time strategies are adopted based on different urgency categories [24]. The higher the urgent score patients have, the shorter time they will wait.
Prior research has empirically investigated the relationships between service utilization, capacity, throughput, and wait times. Service utilization has been shown to have a significant effect on capacity [169], throughput, and wait times in different units (e.g., a congested recovery room and an emergency department) [26, 27, 71]. Capacity has been found to have a positive effect on service utilization, that is, a higher capacity attracts more patients to a hospital, especially non-urgent patients [6, 70]. Capacity has also been discovered to exert a significant negative effect on wait times [27, 71, 170]. Although previous studies have suggested that improvements in throughput often accompany a reduction in wait times [171], the effect of throughput on wait times has not been empirically investigated.

Healthcare units and services have generally evolved in silos focusing on satisfying their own customers [172]. Accordingly, previous research has focused on the relationships between the characteristics within a specific unit. However, we argue that it is inadequate to examine the within-unit relationships in isolation [39, 172], because, in the real world, all the units in a healthcare service system are networked via patient flow. For example, based on the cardiac treatment guidelines [173, 174], units involved in cardiac care are sequentially connected according to patient visits (as shown in Figure 4.2). A directed link between two units denotes that they are temporally related, i.e., patients usually visit the unit the arrow points toward (i.e., the subsequent unit) after visiting the unit the arrow points away from (i.e., the preceding unit). There is usually a “funnel and filter effect” between two temporally related units, as preceding units “determine the absolute numbers of and speed of throughput for patients proceeding” into the subsequent units [38, p.163]. In the context of the CU and the SU, a “diagnostic-therapeutic” cascade effect may also exist, as if more catheterization diagnostic tests are performed, more cardiac surgeries are likely to occur [175, 176, 177]. Thus, investigating the effect of cross-unit relationships, in addition to within-unit relationships, may reveal important insights for wait time management [39].

In summary, the impact factors for a healthcare unit’s performance, wait times
Figure 4.2: The unit framework for cardiac care, drawn from the cardiac treatment guidelines. ECG: Electrocardiogram; PTCA: Percutaneous Transluminal Coronary Angioplasty; PCI: Percutaneous Coronary Intervention.

and throughput, have been studied from the demand and the supply perspectives (as shown in Figure 4.3). The relationships between service utilization, capacity, throughput, and wait times have been investigated within a unit. However, little attention has been paid to the relationships between the characteristics in a cross-unit context, a gap this study aims to fill. We use an SEM-based analysis to explore whether and how the characteristics of one unit exert an influence on the characteristics of other temporally related units, focusing on wait times in particular. Figure 4.3 shows the overall research framework. We choose the CU and the SU as our research context because they both provide key services [24, 165]; they are temporally connected [178]; and published data on the two units are available1.

We use SEM [42, 179] to explore the underlying relationships between the

---

1http://www.ccn.on.ca/
Figure 4.3: The research framework, summarizing the impact factors for throughput and wait times.

characteristics of two units (i.e., CU and SU). Aided by the advantages of SEM in modeling LVs and investigating complex relationships between variables, SEM enables us to identify the complete causal paths of the cross-unit relationships between LVs (i.e., service utilization, capacity, throughput, and wait times), which is not possible using any single traditional statistic method.

The aggregated data for this study was obtained from the CCN\(^1\) and the OPHRDC\(^2\). We choose this data set because it has been collected and published by the Ontario government regularly for more than 10 years. It provides sufficiently comprehensive information on the health care services in Ontario for our study.

\(^2\)https://www.ophrdc.org/Home.aspx
4.2 The Effect of a Unit’s Characteristics on the Wait Times in a Subsequent Unit

Matching fluctuating patient arrivals for healthcare service systems with available capacity is known to be important for improving outcomes such as morbidity and mortality rate, and wait times) [180]. Thus, there has been extensive research examining the relationships between service utilization, capacity, throughput, and wait times, especially within a single unit.

4.2.1 Hypotheses

(1) Within-Unit Relationships

Prior research has shown that service utilization has a positive effect on throughput and wait times. For example, Asaro et al. [71] found that increasing patient arrivals in an emergency department (i.e., service utilization) also increased the throughput and wait times in the department. Schoenmeyr et al. [26] revealed a sensitive relationship between the caseload (i.e., service utilization) and the wait times in a congested recovery room. Harewood et al. [72] found that annual wait times for routine endoscopic procedures lengthened dramatically because of a significant increase in the demand for annual procedures on the endoscopy services. Therefore, we hypothesize that:

Hypothesis 1 (H1): Service utilization has a direct positive effect on throughput within a unit.

Hypothesis 2 (H2): Service utilization has a direct positive effect on wait times within a unit.

In analyzing the current research on the relationship between service utilization and capacity, Baker [169] noted that the desire to meet patient demands was a dominant driving force for capacity changing. Buerhaus [181] pointed out that service utilization increasing for aging population may result in an expanding nursing workforce (human resources) to avoid threatening the healthcare quality.
Justman et al. [182] indicated that HIV scale-up is needed to develop laboratory systems and infrastructures (i.e., physical resources). Several researchers have argued that capacity has a positive effect on service utilization [6, 70]. For instance, Smethurst and Williams [6] noted that for each disease investigated, there were many more patients who did not visit the doctor than there were those who did visit (i.e., “latent” patients). To meet these potential overwhelming patient arrivals, the supplier may increase the system’s capacity. Changes in the capacity may trigger changes in patient arrivals, because more patients are then attracted to that system. However, this argument has not been empirically tested [183]. We thus hypothesize that:

**Hypothesis 3 (H3): Service utilization has a direct positive effect on capacity within a unit.**

Prior research has indicated that capacity is important to ensure better performance in a healthcare service system, measured in throughput and wait times. For instance, Harindra et al. [27] found that supplier capacity was an important factor determining access inequalities (which is usually represented by wait times) in catheterization in Canada. Schoenmeyr et al. [26] showed that the physical capacity of a supplier (e.g., beds) had a significant effect on the wait times in a congested recovery room. Trzeciak and Rivers [170] also found that inpatient capacity (e.g., beds) had an effect on the throughput in an emergency department. Harewood et al. [72] further showed that modifications in routine clinical practice (i.e., service capacity) could significantly affect a procedure’s wait time.

Some studies have revealed that improving capacity may help improve the throughput and the wait times in a healthcare unit. Mukherjee [184] found that improving the management of physicians (e.g., staffing mix) improved patient throughput. Others showed that improving capacity management (such as employing intelligent patient scheduling) shortened wait times efficiently [185, 186]. Therefore, we hypothesize that:
Hypothesis 4 (H4): *Capacity* has a direct positive effect on *throughput* within a unit.

Hypothesis 5 (H5): *Capacity* has a direct positive effect on *wait times* within a unit.

Few studies have investigated the relationship between *throughput* and *wait times*. Brenner et al. suggested that improvements in throughput are often accompanied by a reduction in wait times [171]. An intuitive explanation is that given stable patient arrivals (i.e., determined number of arrivals) in a unit, if resources (physical or human resources) in the unit can be more efficiently used, the patients may be treated quicker. So that the wait times of each patient may be shortened. Therefore, we hypothesize that:

Hypothesis 6 (H6): *Throughput* has a direct negative effect on *wait times* within a unit.

(2) Cross-Unit Relationships

Prior research has examined the relationships of characteristics among several units within a hospital. Alter et al. [38] reported that catheterization has a “funnel and filter” effect on cardiac surgery. Patient arrivals and the capacity of the CU therefore determine the absolute number of and speed of throughput for patients proceeding into the SU. Similarly, prior research has revealed that the CU and the SU have a “diagnostic and therapeutic” cascade effect [175, 176, 177]: if more catheterization diagnostic tests are performed in the CU, more patients may undergo cardiac surgeries. However, these studies do not explain clearly how and to what extent the capacity of one unit may influence the wait times of another. To the best of our knowledge, no prior study has examined whether and to what extent the wait times of one unit influences the wait times of a temporally related unit. We hypothesize that:

Hypothesis 7 (H7): *Service utilization* of the CU has a positive effect on *service utilization* of the SU.
Hypothesis 8 (H8): Capacity of the CU has a positive effect on service utilization of the SU.

Hypothesis 9 (H9): Wait times in the CU have a positive effect on the wait times in the SU.

4.2.2 The Conceptual Model

We postulate a conceptual two-layer wait time model, representing the hypothesized within-unit and cross-unit wait time relationships, as shown in Figure 4.4. The relationships between four characteristics within the CU and the SU are illustrated in Layer 1 and Layer 2. Cross-unit wait time relationships are represented by the effects between the two layers.

Figure 4.4: Conceptual model for this study. Cath: catheterization; Surgery: cardiac surgery; H1-H9: the research hypotheses; +/−: a positive or a negative relationship between two variables.
4.3 SEM Tests and Results

4.3.1 Service Administrative Data in Ontario

The aggregated data used in this study were obtained from the CCN and the OPHRDC in Ontario, Canada. The CCN is a network of 18 member hospitals providing cardiac services in Ontario. The CCN has reported the wait times for the cardiac procedures catheterization, cardiac surgery, and percutaneous coronary intervention in member hospitals across Ontario quarterly since 2004. The reported data include the number of completed cases in a month, the average number of patients waiting at the end of a month, and the monthly average median wait time for each hospital. We are interested in the CU and SU because a regional priority rating score system has been established for these two units (but not other units) in Ontario [167, 168]. The CCN thus provides more detailed statistics for the CU and SU than for other units. Table 4.2 shows the major information provided in the CCN report. We can calculate the variability in the throughput and the wait times for a specific unit from the information in Table 4.2.

We can calculate the monthly average number of arrivals in the CU and SU by:

\[
Arrival_i^t = Throughput_i^t + NoWait_i^t - NoWait_i^{t-1},
\]

(4.1)

where, \(Arrival_i^t\) is the monthly average number of arrivals in quarter \(t\) in unit \(i\), \(Throughput_i^t\) is the monthly average number of patients who have received treatment in quarter \(t\) in unit \(i\), and \(NoWait_i^t\) is the average number of patients waiting at the end of a month in quarter \(t\) in unit \(i\).

The OPHRDC\(^3\), is a definitive source for information on physician use in Ontario. It provides data about physicians in Ontario by specialty (e.g., cardiac surgery, diagnostic radiology) annually. The capacity of SUs is precisely measured by the number of physicians who specialize in cardiac surgery. The capacity of CUs is approximately measured by the number of physicians who perform

\(^3\)https://www.ophrdc.org/Home.aspx
diagnostic radiology, because catheterization is one of the tests that uses radiology, and information about the physicians who perform catheterization is unavailable. However, as the OPHRDC data is grouped by LHIN, not by hospital, the data must be processed to align with the CCN data. Table 4.1 shows the CCN member hospitals and the corresponding LHINs (information retrieved from the CCN\(^4\)). From this table, we can see a direct correspondence between the LHINs and the CCN member hospitals, except for the LHINs of Toronto Central (TC) and North East (NE), each of which has more than one CCN hospital. To facilitate the data analysis, the two LHINs’ data must be decomposed to generate data for the related hospitals.

The data decomposition uses the hospitals’ physician ratio (the number of specific physicians in a hospital to the total number of specific physicians in the corresponding LHIN in 2010) in TC and NE to compute the number of physicians in the relevant hospitals from 2005 to 2008. The physician ratios for the CU and SU in each hospital in TC and NE can be obtained from the website of The College of Physicians and Surgeons of Ontario\(^5\), the governing body for medical doctors in Ontario. After observing the OPHRDC data, we find that in TC and NE, the changes in CUs ranged from 0 to 9 physicians per LHIN year to year (the total average number of catheterization physicians per hospital in the two LHINs is 60), and the changes in SUs ranged from 0 to 1 physicians per LHIN year to year (the total average number of cardiac surgery physicians per hospital in the two LHINs is 7). Therefore, we can assume that the physician ratios in TC and NE are relatively stable, as the physician ratios have been relatively unchanged since 2005. The number of specific physicians in each hospital can therefore be calculated successfully using the specific physician ratio of each hospital multiplied by the number of the specific physicians in the corresponding LHIN each year.

By integrating and processing the two sets of data as discussed above, we obtain

\(^{4}\)http://www.ccn.on.ca/content.php?menuID=14&subMenuID=21&subMenu2ID=14

\(^{5}\)http://www.cpso.on.ca/
Table 4.1: The relationship between the CCN member hospitals and the LHINs

<table>
<thead>
<tr>
<th>LHIN name</th>
<th>CCN member hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td>South West</td>
<td>London Health Sciences Centre</td>
</tr>
<tr>
<td>Waterloo Wellington</td>
<td>St. Mary’s General Hospital</td>
</tr>
<tr>
<td>Hamilton Niagara Haldimand Brant</td>
<td>Hamilton Health Sciences</td>
</tr>
<tr>
<td>Mississauga Halton</td>
<td>Trillium Health Centre</td>
</tr>
<tr>
<td>Toronto Central</td>
<td>Toronto East General Hospitals*</td>
</tr>
<tr>
<td></td>
<td>St. Michael’s Hospital</td>
</tr>
<tr>
<td></td>
<td>University Health Network</td>
</tr>
<tr>
<td></td>
<td>Sunnybrook Health Sciences Centre</td>
</tr>
<tr>
<td>Central</td>
<td>Southlake Regional Health Centre</td>
</tr>
<tr>
<td>South East</td>
<td>Kingston General Hospital</td>
</tr>
<tr>
<td>Champlain</td>
<td>University of Ottawa Heart Institute</td>
</tr>
<tr>
<td>North East</td>
<td>Sault Area Hospital*</td>
</tr>
<tr>
<td></td>
<td>Hôpital Régional de Sudbury Regional Hospital</td>
</tr>
</tbody>
</table>

*: the hospital does not provide cardiac surgery procedures.
comprehensive information about the 11 hospitals (shown in Table 4.2) that provide catheterization and cardiac surgery. Table 4.3 summarizes the characteristics of the two units and their measurements. We focus on the data from 2005 to 2008 (15 quarters in total), because 2004 is the end of the first six-year cardiac expansion plan [102] and the start of the second ten-year cardiac improvement plan [24, 187]. In total, there are 165 data points for CUs and SUs (one quarter in one hospital is regarded as a data point). We describe the statistical analysis methods used to investigate within-unit and cross-unit wait time relationships in the next subsection.

Table 4.2: Cardiac surgery statistics from January 2008 to March 2008, obtained from the CCN

<table>
<thead>
<tr>
<th>Hospital</th>
<th>C</th>
<th>UM (d)</th>
<th>SM (d)</th>
<th>EM (d)</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>London Health Sciences Centre</td>
<td>115</td>
<td>2</td>
<td>5</td>
<td>17</td>
<td>33</td>
</tr>
<tr>
<td>St. Mary’s General Hospital</td>
<td>61</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>Hamilton Health Sciences</td>
<td>127</td>
<td>1</td>
<td>6</td>
<td>12</td>
<td>69</td>
</tr>
<tr>
<td>Trillium Health Centre</td>
<td>79</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>St. Michael’s Hospital</td>
<td>89</td>
<td>5</td>
<td>6</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>Sunnybrook HSC</td>
<td>56</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>University Health Network</td>
<td>129</td>
<td>2</td>
<td>6</td>
<td>13</td>
<td>135</td>
</tr>
<tr>
<td>Southlake Regional HC</td>
<td>75</td>
<td>5</td>
<td>7</td>
<td>28</td>
<td>42</td>
</tr>
<tr>
<td>Kingston General Hospital</td>
<td>47</td>
<td>3</td>
<td>15</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>University of Ottawa Heart Institute</td>
<td>98</td>
<td>6</td>
<td>21</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>Hôpital Régional de Sudbury</td>
<td>36</td>
<td>7</td>
<td>6</td>
<td>19</td>
<td>21</td>
</tr>
</tbody>
</table>

C: the number of completed cases; UM: the median wait time for urgent patients; SM: the median wait time for semi-urgent patients; EM: the median wait time for elective patients; W: the number of patients waiting at the end of a month; d: days.
Table 4.3: A summary of the characteristics of the CU and SU

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Measurements</th>
<th>CU</th>
<th>SU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service utilization</td>
<td>Monthly number of arrivals</td>
<td>340</td>
<td>82</td>
</tr>
<tr>
<td>Capacity</td>
<td>Number of physicians, yearly</td>
<td>60</td>
<td>7</td>
</tr>
<tr>
<td>Throughput</td>
<td>Monthly number of completed patients</td>
<td>346</td>
<td>83</td>
</tr>
<tr>
<td>Wait times</td>
<td>Median wait time of U/S/E patients</td>
<td>1/10/15</td>
<td>3/6/19</td>
</tr>
<tr>
<td></td>
<td>Number of waits at the end of a month</td>
<td>101</td>
<td>58</td>
</tr>
</tbody>
</table>

CU: Catheterization unit; SU: Cardiac surgery unit; U: urgent; S: semi-urgent; E: elective.

4.3.2 SEM Tests

We use PLS-based SEM [43] to test the proposed two-layer wait time model (shown in Figure 4.4) and the related hypotheses as this study is exploratory rather than confirmatory. The software SmartPLS\(^6\) is used for path modeling and PLS-based data analysis.

In SEM-based data analysis, the measurements for wait times are modeled as formative indicators rather than reflective ones [43, 188]. A formative model is used when a latent construct (i.e., a factor, such as service utilization, capacity, throughput, and wait times) is viewed as an explanatory combination of its manifest variables (i.e., measurements) [144, 189]. In contrast, in a reflective model, the manifest variables are viewed as being caused by a underlying common dimension or a construct [189]. Here, the manifest variables for wait times are not interchangeable or correlated with one another because they measure wait times from different perspectives. Therefore, the LV wait times is the summation of its corresponding manifest variables. In other words, the measurement items of wait times will be formative in the construct of wait times.

We use the data for the CU and SU in the same quarter to test the cross-unit

\(^6\)http://www.smartpls.de/
relationships. As the longest wait time for a patient in the CU is around one month, we can assume that the great majority of patients who need cardiac surgery will be transferred from the CU to the SU within a quarter. In the next section, we present the results of the PLS analysis, focusing on how the characteristics affect one another within a unit and how the characteristics of the CU affect the characteristics of the SU, particularly the SU’s wait times.

4.3.3 Test Results

(1) Within-Unit Relationships

As illustrated in Figure 4.5, in support of H1-H3, service utilization has a significant positive effect on throughput, capacity, and wait times. The path coefficients for the effect of service utilization on throughput are $\beta = 0.585$ ($t=18.677, p < 0.01$) for the CU and $\beta = 0.797$ ($t=35.115, p < 0.01$) for the SU. The path coefficients for the effect of service utilization on capacity are $\beta = 0.921$ ($t=127.754, p < 0.01$) for the CU and $\beta = 0.574$ ($t=25.219, p < 0.01$) for the SU. The path coefficients for the effect of service utilization on wait times are $\beta = 0.619$ ($t=2.908, p < 0.05$) for the CU and $\beta = 0.472$ ($t=6.111, p < 0.01$) for the SU. These results confirm findings from prior studies [26, 27, 71, 72, 169], providing further evidence that service utilization is an important predictor for capacity, throughput and wait times within a healthcare unit.

In support of H4, capacity has a significant positive effect on throughput. The path coefficients for the effect of capacity on throughput are $\beta = 0.410$ ($t=13.162, p < 0.01$) for the CU and $\beta = 0.155$ ($t=5.914, p < 0.01$) for the SU. These results also confirm findings from prior studies [170, 184], suggesting that an improvement in capacity will lead to improved throughput within a unit.

Hypothesis H5 is only partially supported by our data. For the CU, capacity has a significant negative effect on wait times ($\beta = -0.252, t=2.465, p < 0.01$), thus supporting H5. However, for the SU, capacity has a significant positive effect on
Figure 4.5: PLS test results based on a formative measurement model. Cath: catheterization; Surgery: cardiac surgery.

*wait times* ($\beta = 0.115, t=3.071, p < 0.01$), which does not support H5. This finding differs with prior studies [26, 72], which suggests that an improvement in a unit’s capacity can significantly shorten its patients’ wait times.

The positive effect of *capacity* on *wait times* for the SU can be explained using Smethurst and Williams’s work [6, 70]. They found that hospital waiting lists are self-regulating. When capacity increases to meet the needs of patients, the number of patient arrivals may change again, creating an even greater number of patient arrivals. A mass of “hidden” patients [6] who have diseases but have not been willing to go to a hospital, may be persuaded to visit that hospital if they believe that they will be treated quicker. Hence, expanding the capacity of the SU may help shorten wait times temporarily, but the wait times will then increase beyond
their initial values, as patient arrivals increase in response to the larger capacity.

Hypothesis H6 is not supported by our data. Throughput has a significant positive effect on wait times ($\beta = 0.352, t=1.659, p < 0.1$) in the CU, whereas the effect of throughput on wait times is negligible for the SU ($\beta = 0.049, t=0.593, p > 0.1$). This finding suggests that throughput and wait times have similar changing patterns in the CU, but not in the SU, which is contrary to the expectation that an improvement in the throughput will result in an improvement in wait times.

Urgent patients’ queue jumping behavior may explain the positive relationship between throughput and wait times in the CU. Queue jumping means that urgent patients can skip the queue and jump to any position on a waiting list because of their treatment priority [190]. If more urgent patients arrive, units delay the treatment of the semi-urgent and elective patients to serve the high priority patients in time, indirectly making the non-urgent patients wait longer. The overall wait times for the unit may increase as a result. The absence of a significant relationship between the throughput and wait times in the SU could be because the SU has much fewer urgent patients than the CU does. For instance, in the fiscal year of 2004, the percentage of urgent patients in the CU in Ontario was 49% (out of a total of 52628 patients), whereas the percentage of urgent patients in the SU was only 23% (out of a total of 7825 patients) [24]. This finding implies that, in some cases, throughput and wait times may not be directly related to reflect the quality of a unit’s performance.

(2) Cross-Unit Relationships

As show in Figure 4.5, H7 is not supported by our data ($\beta = 0.022, t = 0.277, p > 0.1$). The service utilization of the CU does not have a significant effect on the service utilization of the SU. The capacity of the CU has a significant positive effect on the service utilization of the SU ($\beta=0.644, t = 8.498, p < 0.01$), which supports H8.

These two findings explain the formation of the “funnel and filter” effect [38] between the CU and the SU. They suggest that more arrivals in the CU usually
lengthen the waiting list, but do not heavily affect the throughput proceeding to the SU. In reality, the CU always has a waiting list, as can be seen in the CCN data. However, the capacity of the CU heavily determines the absolute number of and speed of throughput for patients proceeding into the SU, forming the “funnel and filter”.

In support of H9, the results of our analysis reveal that the wait times in the CU have a significant positive effect on the wait times in the SU ($\beta = 0.330$, $t = 9.859$, $p < 0.01$). This is strong evidence that the wait times in the CU are an important predictor for the wait times in the SU. A possible explanation for this effect is a delay cascade [191]. Unnikrishnan et al. [191] simulated and observed that delays will cascade in an emergency department network. In that study, all of the emergency departments in different hospitals were networked by the transfer paths of ambulances. In other words, delays in an emergency department will result in the wait times increasing in other emergency departments nearby. Cardiac care has a similar unit network (shown in Figure 4.2) in a hospital. Therefore, delays in one unit may spread to other related units in the unit network, forming the direct cross-unit wait time relationship.

Table 4.4 summarizes the hypothesis testing results. An examination of our results, shown in Figure 4.5, reveals both direct and indirect causal paths from the characteristics of the CU to the wait times in the SU. The service utilization and capacity of the CU also have indirect effect on the wait times in the SU, in addition to the direct effects. In other words, the wait times in the SU may be influenced by the CU via the following causal paths: wait times in the CU $\rightarrow$ wait times in the SU; service utilization of the CU $\rightarrow$ capacity of the CU $\rightarrow$ service utilization of the SU $\rightarrow$ wait times in the SU; and service utilization of the CU $\rightarrow$ capacity of the CU $\rightarrow$ service utilization of the SU $\rightarrow$ capacity of the SU $\rightarrow$ wait times in the SU. The service utilization of the CU appears to be the most essential driving force for the wait time dynamics in the CU and in the SU.
Table 4.4: Hypothesis testing results

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Supported?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1-H4, H8, H9</td>
<td>Fully supported</td>
</tr>
<tr>
<td>H5</td>
<td>Partially supported</td>
</tr>
<tr>
<td>H6, H7</td>
<td>Not supported</td>
</tr>
</tbody>
</table>

4.4 Discussion

In this chapter, we have examined whether and how the characteristics of a preceding unit can affect the *wait times* in the SU. Unlike prior studies, we used SEM to assess the cross-unit wait time relationships from data published on healthcare services in Ontario, Canada. The results of our analysis validate the proposed conceptual model, thus providing empirical support for the hypothesized relationships between the characteristics *service utilization*, *capacity*, *throughput*, and *wait times*, both within a unit and across units.

Our results show that the *wait times* in the CU have a direct positive effect on the *wait times* in the SU. This is a novel result, as prior research has seldom examined the influence of one unit’s *wait times* on the *wait times* in a subsequent unit in the patient flow process. A possible explanation for the effect is a delay cascade in the cardiac care unit network (Figure 4.2), proposed by Unnikrishnan et al. [191].

The results of our analysis provide empirical evidence for previous findings that within a unit, *service utilization* has a positive effect on *capacity*, *throughput*, and *wait times*; within a unit, *capacity* has a positive effect on *throughput*; and that across units, the *service utilization* of one unit will be positively influenced by the *capacity* of the preceding unit.

We also obtained the surprising findings that the relationship between *capacity* and *wait times* differs in units with different profiles (e.g., different patient proportion in each urgency category); *throughput* has a positive effect on the *wait*
times in a unit; there are direct and indirect wait time relationships between temporally-related units; and that service utilization of the CU is an essential predictor for the other characteristics of the CU and SU.

However, there may be other factors affecting a unit’s performance in addition to service utilization, capacity, and cross-unit relationships. For example, the patient risk profile (i.e., the value of predicted operative mortality) has been identified as a factor that may affect triage or referral patterns and the allocation of resources [192]. Although the exact effects of patient risk profiles on a health care system’s performance (wait times in particular) are still unclear, these relationships should be explored by incorporating patient risk into our conceptual model.

There are different methods for calculating the value of risk for patients undergoing catheterization (e.g., SYNTAX7) and cardiac surgery (e.g., EuroSCORE8 and Higgins Score [193]) based on several risk factors. For example, the surgical risk factors for isolated coronary artery bypass graft (CABG) surgery include age, sex, previous CABG, left ventricular function, and coronary anatomy, among others. [178, 194]. The Institute for Clinical Evaluative Science of Ontario has published data on the distribution of risk profiles in isolated CABG, the major type of cardiac surgery, in 2005 and 2006 in Ontario hospitals [178]. We used this published risk profile data (represented by the percentage of low-, medium-, and high-risk patients for catheterization in a hospital), to investigate the relationship between risk profiles and wait times. The missing data for each hospital’s risk profiles for 2007 and 2008 are substituted with the mean value of the available risk data for that hospital [178], which is a common method for handling missing data in statistical data analysis [195, 196]. By integrating our original cardiac care data with the risk profile data, we conduct an additional PLS analysis to test the extended two-layer wait time model, with risk profiles added as an extra predictor of wait times in the SU (see Figures 4.6, 4.7, and 4.8).

The results of the analysis (Figures 4.6, 4.7, and 4.8) reveal that the pattern of

---

7http://www.syntaxscore.com/
8http://www.euroscore.org/
within- and cross-unit relationships (i.e., hypotheses H1-H9) between characteristics (i.e., service utilization, capacity, throughput, and wait times in the CU and SU) remain unchanged. When risk profiles are represented differently, as a percentage of low-risk patients, percentage of medium-risk patients, or percentage of high-risk patients, they can have different effects on the wait times in the SU.

The percentage of low-risk patients has a significant negative effect on wait times (see Figure 4.6). The explanation for this finding is still unclear as almost no prior work has addressed this issue to the best of our knowledge. However, we postulate that the treatment process for low-risk patients is easier than for higher-risk patients, and hence, the length of stay (including the pre-operative, operating, and post-operative stay) of low-risk patients may be shorter than higher-risk patients. Therefore, if there are more low-risk patients in the SU, the total wait times in this unit will decrease.

Interestingly, the percentage of medium-risk patients has a significant positive effect on wait times (see Figure 4.7). This may be due to unexpected upgrading of the patients proceeding to cardiac surgery to more urgent categories (e.g., upgrading the medium-risk patients from semi-urgent to urgent) [197, 198]. The upgrading event may trigger queue jumping behavior [190], which will hinder the normal treatment schedule and result in longer wait times. This observation is consistent with the prior findings that proportionately more patients in the more urgent categories than in the less urgent categories may have wait times in excess of the maximum acceptable [199].

The percentage of high-risk patients does not have a significant effect on wait times (see Figure 4.8), contrary to our expectation. Prior work indicates that high-risk patients tend to be assigned higher priorities in the triage process [197], and thus more high-risk patients may imply more urgent patients. As urgent patients are more likely to undergo expedited surgery, treatment for non-urgent patients may be delayed, resulting in prolonged overall wait times [190]. Although, we do
not yet have a sound explanation for this unexpected lack of effect, the observed inconsistency between the effect of the high-risk profile and that of the medium-risk profile may be due to the methodology used to stratify the patient risk profiles and priority categories, an issue that deserves further investigation.

4.5 Summary

In this chapter, we used SEM-based analysis to examine whether and how the characteristics of a preceding unit exert effects on the wait times in subsequent units, focusing on cardiac care services. Unlike existing studies, we used SEM to explore the within-unit and cross-unit wait time relationships from aggregated
Figure 4.7: PLS test results for the extended two-layer wait time model with the medium-risk profile in the SU. Cath: catheterization; Surgery: cardiac surgery.

data. We investigated two temporally related units in cardiac care, the CU and its subsequent unit, the SU. Our results reveal that wait times in the CU have a direct positive effect on wait time in the SU; capacity of the CU has a direct positive effect on service utilization of the SU; within each unit, there are significant relationships between the characteristics, except for the effect of throughput on wait times in the SU; different patient risk profiles may affect wait times in the SU in different ways (e.g., positive or negative effects). The findings presented in this chapter suggest that when healthcare administrators seek to alleviate wait times in a health care system, they should consider the cross-unit wait time relationships and take into account the relationship between priority triage and risk stratification, especially for cardiac surgery. This chapter is an additional example
Figure 4.8: PLS test results for the extended two-layer wait time model with the high-risk profile in the SU. Cath: catheterization; Surgery: cardiac surgery.

demonstrating the effectiveness of an SEM-based analysis for discovering the complex relationships between multiple observed variables and LVs.
Chapter 5

Projecting the Changes of Service Performance

The previous two chapters presented the use of an SEM-based analysis to discover the complex effects of specific factors on service utilization or wait times. This chapter further addresses the problem of how to project the changes of healthcare service performance with respect to demographic shifts, which is one of the major concerns for better planning and allocating of future healthcare resources. We propose an integrated projection method that consists of the following steps: applying an SEM-based analysis to identify the complex relationships between demographic profiles and healthcare service characteristics (e.g., capacity, supply, utilization, and performance); carrying out projection to estimate service utilization and service performance based on the discovered complex relationships and demographic shifts; conducting queueing model analysis to gain insights into the changing patterns of the estimated service performance over time. We then use the proposed method to estimate the changes in service utilization and performance of cardiac surgery services in Ontario, Canada. Figure 5.1 shows the research focus in this chapter with respect to the larger context of understanding a healthcare service system.
Figure 5.1: A schematic diagram illustrating the use of integrated projection to project the changes in service performance with respect to demographic shifts. (a) The research focus of this chapter (highlighted in red) is part of the larger context of understanding a healthcare service system. (b) Carrying out an estimation based on integrated projection.

5.1 Introduction

Many areas in the world now face notable demographic changes/shifts due to aging and immigration [37, 200]. For example, it was projected that the population aged 65 and above in Ontario, Canada, would increase from 1.8 million in 2010 to 4.1 million in 2036, accounting for 13.9% and 23.4% of the total population, respectively [36]. The number of new immigrants in Ontario was projected to increase by 0.107 to 0.135 million annually from 2010 to 2036, accounting for nearly 70% of the total population growth [36].

Demographic shifts (e.g., age and ethnic profiles) are known to have a direct effect on healthcare service utilization due to their correlations with risk factors
for certain diseases and with service utilization behavior. For example, risk factors associated with cardiovascular diseases are more prevalent in the population aged 50 years old and above [14, 22]. Ethnic groups differ in their risks for cardiovascular diseases [22, 45, 201] and in their healthcare service utilization behavior [202, 203].

Demographic shifts will also have an effect on the performance (i.e., throughput and wait times) of a healthcare service. It has been found that healthcare service performance is affected not only by supply factors, such as physical and human resources, and management strategies [47, 204], but also by the dynamics of patient arrivals in terms of volume and characteristics (e.g., patient profile and severity of diseases with various co-morbidities) [29, 204]. An in-depth understanding of the potential changes in healthcare service characteristics (e.g., service utilization and performance) due to demographic changes will be helpful for middle-/long-term healthcare resource planning and allocation.

We address three research questions.

- The relationships between demographic profiles and healthcare service utilization involve several factors with direct and/or indirect, linear and/or nonlinear, and dynamic interactions. How can we learn these multi-factor complex relationships from limited aggregated data?

- Once we have found the multi-factor complex relationships between demographic profiles and healthcare service utilization, how can we project the changes in service utilization with respect to demographic shifts?

- Estimation results based on multi-factor complex relationships are somewhat uncertain and cannot demonstrate the dynamics of estimated service utilization over time. How can we determine the dynamically changing process of healthcare service utilization with respect to demographic shifts?
Figure 5.2: A schematic diagram of the three-step integrated projection method and its application to cardiac surgery services.
To answer these questions, we propose a method of *integrated projection*, which uses an *SEM-based analysis* to discover the complex effects of multiple factors on service utilization and service performance, carries out *projections* to estimate healthcare service utilization based on the derived multi-factor complex relationships, and constructs a *queueing model* to simulate the dynamics of the estimated performance over time.

We apply the integrated projection method to estimate the changes in service utilization and service performance in cardiac surgery services in Ontario, Canada. Our method is shown to be able to identify the complex relationships between the age profile, recent immigrant profile, and characteristics of cardiac surgery; describe the variations in healthcare service utilization with respect to demographic shifts; and demonstrate the temporal changes in estimated cardiac surgery performance using queueing model simulations.

### 5.2 Integrated Projection

We propose an analytical method (shown in Figure 5.2) to unveil the underlying relationships between demographic shifts and healthcare service utilization.

1. *Analysis of complex relationships between factors*: Based on a training data set (e.g., statistics), we use SEM [42] to identify the complex relationships between multiple factors (e.g., the age and recent immigrant profiles, and cardiac surgery characteristics in our case study).

2. *Qualitative projection*: Based on the identified multi-factor complex relationships from the first step, we propose a set of equations to estimate the changes in healthcare service utilization (e.g., cardiac surgery utilization in our case study) with respect to demographic shifts.

3. *Dynamics simulation*: Based on the estimated service utilization, we build specific queueing models to simulate the operation of different healthcare
systems, such as cardiac surgery ORs (CS-ORs) in our case study, to gain insights into the performance dynamics of their provided services over time.

5.2.1 SEM-Based Analysis

SEM uses a measurement model and a structural model to explore the complex relationships between factors/variables (as shown in Figure 5.3). The measurement model [42] characterizes the linear relationships between observed variables (also known as MVs) and the corresponding LVs. One of the typical ways to relate MVs to LVs is through the reflective measurement model, in which each LV is reflected in its corresponding MV. Formally, let $\Xi = \{\xi_1, \xi_2, \ldots, \xi_N\}$ be a set of LVs and $X_{\xi_j} = \{x_{j1}, x_{j2}, \ldots, x_{N_\xi M_{\xi_j}}\}$ be a set of MVs relating to $\xi_j$ ($\forall j \in [1, N_\xi]$), where $N_\xi = |\Xi|$ denotes the total number of LVs and $M_{\xi_j} = |X_{\xi_j}|$ denotes the total number of MVs which relate to $\xi_j$ ($\forall j \in [1, N_\xi]$). The relationship between $x_{jk}$ ($\forall k \in [1, M_{\xi_j}]$) and its related $\xi_j$ can be expressed as follows [205]:

$$x_{jk} = \pi_{jk0} + \pi_{jk} \xi_j + \varepsilon_{jk},$$  \hspace{1cm} (5.1)

where $\pi_{jk0}$ and $\pi_{jk}$ (i.e., loading in SEM) are the regression parameters and $\varepsilon_{jk}$ is the residual error.

![Figure 5.3: The basic components in SEM.](image)

The structural model [42] describes the linear relationships between the LVs. Formally, let $\tilde{\Xi}_{\xi_j}$ ($\tilde{\Xi}_{\xi_j} \subset \Xi$) be a set of LVs that $\xi_j$ relates to. The relationships
between \( \xi_j \) (\( \xi_j \in \Xi \)) and its related LVs (\( \tilde{\Xi}_{\xi_j} \)) can be written as follows [205]:

\[
\xi_j = \beta_{j0} + \sum_i \beta_{ji} \xi_i + \zeta_j,
\]

(5.2)

where \( \beta_{j0} \) is a constant number, \( \beta_{ji} \) (i.e., the path coefficient in SEM) is the regression weight of \( \xi_i \) (\( \forall \xi_i \in \tilde{\Xi}_{\xi_j} \)) relating to \( \xi_j \), and \( \zeta_j \) is the residual error.

We test multi-factor complex relationships using PLS-based SEM, as it is more suitable for exploratory studies, as in our case study, than covariance-based SEM [43].

### 5.2.2 Projection

In this subsection, we describe how to estimate healthcare service utilization based on demographic shifts and multi-factor complex relationships. The estimation process follows four sub-steps.

**S1:** Calculate the value of any exogenous LV with respect to the change in each corresponding MV using Equation 5.3. An exogenous LV \( \xi_j \) (\( \xi_j \in \Xi \)) is an LV that does not vary due to other LVs.

\[
\xi'_j = f(\theta_{jk}|\pi_{jk}, \sigma_{jk}, \Delta, x_{jk}) = \frac{x_{jk}(1 + \Delta, x_{jk})^\tau}{\pi_{jk}} - \sigma_{jk} + \theta_{jk},
\]

(5.3)

where \( \xi'_j \) is the estimation value of \( \xi_j \) given the changes in its MV, \( x_{jk} \) (\( x_{jk} \in X_{\xi_j} \)), in the estimation time \( \tau \); \( \sigma_{jk} = \frac{\pi_{jk0} + \varepsilon_{jk}}{\pi_{jk}} \) represents a constant value; \( \Delta x_{jk} \) is the changing rate of \( x_{jk} \) per time unit; and \( \theta_{jk} \) represents how \( \xi_j \) will change in accordance with a variation in \( x_{jk} \).

**S2:** Take an estimation value for each exogenous LV based on Equation 5.4. Let \( X'_{\xi_j} \) be a set of changed MVs related to an exogenous LV, \( \xi_j \) (\( X'_{\xi_j} \subseteq X_{\xi_j} \)), and let \( M'_{\xi_j} = |X'_{\xi_j}| \) denote the number of MVs that change in the estimation time \( \tau \). As each exogenous LV, \( \xi_j \), has \( |X'_{\xi_j}| \) estimated values in accordance with the changes in the MVs \( X'_{\xi_j} \), we minimize the expectation of \( \theta_{jk} \) (\( k \in [1, M'_{\xi_j}] \)).
to get one reasonable estimation value for the exogenous LV, $\xi_j$.

$$
\xi'_j = \arg \min_{k \in [1, M'_j]} E(f(\theta_{jk}|\pi_{jk}, \sigma_{jk}, \Delta x_{jk})).
$$  \hfill (5.4)

It should be noted that if all of the MVs related to an exogenous LV, $\xi_j$, do not change during the estimation time $\tau$, then the estimation value of $\xi_j$, i.e., $\xi'_j$, will be equal to the original value of $\xi_j$ discovered by SEM.

**S3:** Calculate any endogenous LV, $\xi_j$ ($\xi_j \in \Xi$), with Equation 5.5, based on the multi-factor complex relationships learned by SEM. An endogenous LV is an LV which varies depending on other LVs.

$$
\xi'_j = \beta_{j0} + \sum_i \beta_{ji}\xi'_i + \zeta_j.
$$  \hfill (5.5)

where $\xi'_j$ is the estimated value of $\xi_j$ given the estimated values of its related LVs, $\tilde{\Xi}_{\xi_j}$.

**S4:** Calculate the MVs related to each endogenous LV using Equation 5.1.

### 5.2.3 Queueing Model Simulation

Queueing models are useful for simulating the operation of healthcare systems and investigating interrelated processes, such as arriving at a queue and waiting [50, 96, 26]. A general queueing model (shown in Figure 5.4) for simulating a healthcare service system should define four basic characteristics.

- **Patient types and arrival patterns** (commonly denoted by $\lambda$): Patients in a healthcare service can be divided into different types according to their characteristics. For example, as shown in Figure 5.4, patients are usually categorized as urgent, semi-urgent, and elective. Different patient groups may differ in arrival rates and received services, such as service priority and service time. The arrival pattern is usually represented by a statistical distribution of inter-arrival times.
Figure 5.4: An schematic diagram of a general queueing model for healthcare. E: elective; S: semi-urgent; U: urgent.

- **Patient behavior**: Some patients may be sensitive to wait times and may quit a queue if they have to wait too long, whereas others may be willing to stay in a queue no matter how long they will wait. This patient behavior may affect the performance of a healthcare service system and thus should be stated clearly in modeling.

- **Service capacity and service patterns** (commonly denoted by $\mu$): In a healthcare service system, the service capacity usually corresponds to the number of service stations (e.g., the number of ORs in our case study). Patients' service times follow a specific distribution (e.g., an exponential distribution or a uniform distribution).

- **Service discipline**: Service discipline determines the order that patients in a queue are served in after they arrival. Some commonly used service disciplines in a healthcare service system are first come first served (e.g., outpatient service), and priority-based, as in our case study.
5.3 Estimating the Performance of Cardiac Surgery Services

We apply the proposed integrated projection method to the cardiac surgery services in Ontario, Canada to discover the effects of demographic profiles on cardiac surgery utilization and performance; project how cardiac surgery utilization and performance change in response to demographic shifts; and demonstrate the dynamics of cardiac surgery performance in terms of queue length and wait times by modeling and simulating the operational process of CS-ORs. The analytical process in this case study is illustrated in Figure 5.2.

We introduce the aggregated data set used in Section 5.3.1. In Section 5.3.2, we describe the hypothetical complex relationships between the age profile, recent immigrant profile, cardiac surgery capacity (i.e., the number of CS-ORs), supply (i.e., the number of physicians able to perform cardiac surgeries), service utilization (i.e., the number of patient arrivals), and performance (i.e., the throughput and wait times of semi-urgent/elective patients) and test these relationships with SEM. The process of projecting cardiac surgery utilization based on SEM test results is shown in Section 5.3.3. The multi-server multi-queue with an entrance control queueing model (MSMQ-EC) used to simulate CS-ORs is shown in Section 5.3.4. We present and discuss the estimation and simulation results in Section 5.3.5.

5.3.1 Demographic and Service Administrative Data in Ontario

We apply our proposed method to the healthcare service system in Ontario, Canada, which contains 14 LHIN\(^1\) areas and 11 hospitals providing cardiac surgery services. Aggregated data describing cardiac surgery characteristics was obtained primarily from the CCN\(^2\). Reports from the CCN and OPHRDC\(^3\) provided the aggregated data

---

\(^1\)http://www.lhins.on.ca/home.aspx  
\(^2\)http://www.ccn.on.ca/  
\(^3\)https://www.ophrdc.org/Home.aspx
on the number of ORs [102] and physicians for cardiac surgery in each hospital.

Demographic information on patient age and recent immigration in each LHIN were obtained from Statistics Canada\(^4\). We selected 47 major cities and towns in Ontario with a population of more than 40,000 to represent the demographic profiles of the LHINs.

The age profile in an LHIN is defined as the ratio of the population aged 50 years and above to the total population in the LHIN. This age population is of interest because it is the major cohort of cardiac surgery patients [134]. An LHIN’s recent immigrant (RI) profile is the ratio of recent immigrants from Asia and Africa to the total population in the LHIN. We are interested in these ethnic groups because they account for approximately 70% of new immigrants and prior work has shown that, compared to white groups, the risk factors for cardiovascular diseases are more prevalent in black, South Asian, Southeast Asian, West Asian and Middle Eastern ethnic groups [22, 45]. As patients dwelling in one LHIN may travel to other LHINs to receive cardiac surgeries, we preprocess the demographic data by the cross-LHIN ratio reported by the CCN, so as to more precisely characterize the demographic profiles for each LHIN.

As shown in Figure 5.2, we use the aggregated data concerning demographic profiles and cardiac surgery characteristics from 2005 to 2007 (12 quarters) as the training data set. The aggregated data from 2008 to 2012 was used to calculate the changes in the demographic profiles and to evaluate the estimated cardiac surgery service utilization and performance. Table 5.1 provides an overview of the training data set.

\[5.3.2\] Relationships between Demographic Factors and Service Characteristics

We must first derive the hypothetical complex relationships between demographic profiles, cardiac surgery service utilization, and service performance, before the

\(^4\)http://www.ices.on.ca/
### Table 5.1: The training data set

<table>
<thead>
<tr>
<th>LV</th>
<th>MV</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age profile</td>
<td>Ratio of population aged 50 and over in an LHIN</td>
<td>0.34</td>
</tr>
<tr>
<td>RI profile</td>
<td>Ratio of recent immigrants from Asia and Africa in an LHIN</td>
<td>0.05</td>
</tr>
<tr>
<td>Service utilization</td>
<td>Average number of patient arrivals, monthly</td>
<td>82</td>
</tr>
<tr>
<td>Capacity</td>
<td>Number of cardiac surgery physicians, yearly</td>
<td>7</td>
</tr>
<tr>
<td>Supply</td>
<td>Number of CS-ORs</td>
<td>3</td>
</tr>
<tr>
<td>Throughput</td>
<td>Average number of completed patients, monthly</td>
<td>83</td>
</tr>
<tr>
<td>Wait times</td>
<td>Median wait time of S/E patient*</td>
<td>6/19</td>
</tr>
<tr>
<td></td>
<td>Queue length, monthly</td>
<td>58</td>
</tr>
</tbody>
</table>

*RI profile: recent immigrant profile; S: semi-urgent patient; E: elective patient; *: the median wait time of urgent patient is not considered a measurement for the LV wait times because it does not significantly reflect wait times according to our pre-data analysis.*

Relationships can be tested with SEM. According to prior work, aging and immigration are two major factors accounting for demographic changes [37, 200]. Both age profile [14, 22] and RI profile [22, 45, 201] may have a positive effect on cardiac surgery service utilization. Cardiac surgery capacity and supply may have effects on service utilization [6, 206], throughput [29], and wait times [29]. Throughput may have an effect on wait times [29]. We then use the hypothetical relationships to empirically examine the effects of the demographic profiles on cardiac surgery characteristics, as shown in Figure 5.5. The MVs for each LV in Figure 5.5 are listed in Table 5.1.
Figure 5.5: The hypothetical relationships between the demographic profiles, service utilization, and service performance.

### 5.3.3 Service Performance Projection

Once the relationships between the demographic factors and service characteristics have been identified, cardiac surgery performance in response to demographic changes can be projected.

**S1:** Calculate the values of the exogenous LVs, i.e., the age profile, RI profile, capacity, and supply. As the cardiac surgery capacity and supply only change slightly from one year to the next, according to real-world observations, we assume that the cardiac surgery capacity and supply will not change during the estimation time. We can therefore make a clearer observation on how demographic shifts affect cardiac surgery service utilization and performance, and whether existing cardiac surgery resources are capable of providing a stable service in terms of wait times. The changes in the age and RI profiles can be expressed by:

\[
\begin{align*}
\xi_1 &= x_1(1 + \Delta Age)^\tau, \\
\xi_2 &= x_2(1 + \Delta RI)^\tau,
\end{align*}
\]

where \(\Delta Age\) and \(\Delta RI\) are the changes in the age and RI profiles, respectively, at time \(\tau\).
S2: Calculate the values of the endogenous LVs, i.e., service utilization, throughput, and wait times, using:

\[
\begin{align*}
\xi_5 &= \beta_{50} + \beta_{15}\xi_1 + \beta_{25}\xi_2 + \beta_{35}\xi_3 + \beta_{45}\xi_4 + \xi_5 \\
\xi_6 &= \beta_{60} + \beta_{36}\xi_3 + \beta_{46}\xi_4 + \beta_{56}\xi_5 + \xi_6 \\
\xi_7 &= \beta_{70} + \beta_{37}\xi_3 + \beta_{47}\xi_4 + \beta_{57}\xi_5 + \beta_{67}\xi_6 + \xi_7.
\end{align*}
\] (5.7)

S3: Calculate the values of the MVs that relate to the endogenous LVs, i.e., the number of patient arrivals, queue length, and median wait times for semi-urgent/elective patient. We can therefore estimate the values of cardiac surgery service utilization and performance with respect to changes in the demographic profiles.

5.3.4 The MSMQ-EC Queueing Model

We build an MSMQ-EC queueing model based on the real-life execution of CS-ORs in Ontario, to gain insights into the temporally changing patterns in cardiac surgery performance with respect to demographic shifts. The MSMQ-EC is similar to the queueing model in our prior work [207]:

- \( M \) homogeneous ORs with the same service rate \( \mu \);
- Three patient groups, urgent (U), semi-urgent (S), and elective (E), with the arrival rates \( \lambda_U \), \( \lambda_S \), and \( \lambda_E \), respectively;
- \( N \) physicians maintaining \( N \) priority queues.

The service principle of the queueing model is as follows. Urgent patients have the highest priority and should be immediately settled in an available CS-OR. If all of the CS-ORs are occupied, urgent patients must wait and bump the first available CS-OR block for non-urgent patients. Semi-urgent and elective patients are scheduled by physicians following a priority-based service principle. A new incoming non-urgent (\( \bar{U} \), i.e., semi-urgent or elective) patient will first be assigned to a physician \( j \) (\( \forall j \in [1, N] \)) with a probability \( p_{j,\bar{U}} \). Physician \( j \) performs a non-urgent surgery with a
probability \( q_{j,U} \) that represents the “entrance control” of CS-ORs for non-urgent surgeries. Similar to our prior work [207], \( p_{j,U} \) and \( q_{j,U} \) follow uniform distributions in the simulation.

### 5.3.5 Projection Results

In this subsection, we demonstrate the estimation results of cardiac surgery utilization in the Hamilton Health Science Centre (HHSC) hospital, located in the Hamilton Niagara Haldimand Brant LHIN (LHIN 4), between 2008 and 2011. The complex relationships in cardiac surgery are extracted from the training data set using the software SmartPLS\(^5\). The simulation results based on the MSMQ-EC queueing model demonstrate the dynamics of cardiac surgery performance. In the simulation, the queueing model was parameterized using the general operational data from the HHSC CS-ORs in 2007. All of the simulation studies are implemented using the discrete-event simulation toolbox SimEvents in MATLAB 2010.

**1) The Results of SEM Tests and Service Utilization Estimation**

According to the PLS test results, shown in Figure 5.6, service utilization has an \( R^2 \) of 0.631, throughput has an \( R^2 \) of 0.874, and wait times has an \( R^2 \) of 0.610. These endogenous LVs are therefore well explained by their dependent variables. For example, the \( R^2 \) of the LV service utilization reflects that its dependent variables, i.e., age profile, RI profile, capacity, and supply, explain 63.1\% of the variance in service utilization. Although there may be other factors influencing service utilization other than the impact factors we have considered, the results imply that the factors used capture most of the variation in service utilization. The PLS test results show that all of the hypothetical effects between the LVs are significant, except the effects of supply on throughput and wait times, and the effect of throughput on wait times.

It should be noted that LVs with significant correlations (shown in Figure 5.6 (a)) can be considered in the projection calculation process. The path coefficients

\(^5\)http://www.smartpls.de/
Table 5.2: The estimated values for cardiac surgery utilization and performance (average value in a month)

<table>
<thead>
<tr>
<th></th>
<th>2010&lt;sub&gt;E&lt;/sub&gt;</th>
<th>2011&lt;sub&gt;E&lt;/sub&gt;</th>
<th>2010&lt;sub&gt;A&lt;/sub&gt;</th>
<th>2011&lt;sub&gt;A&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service utilization</strong></td>
<td>108</td>
<td>115</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Throughput</strong></td>
<td>98</td>
<td>102</td>
<td>83</td>
<td>84</td>
</tr>
<tr>
<td><strong>Wait times</strong></td>
<td>14.53</td>
<td>14.98</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Queue length</strong></td>
<td>65</td>
<td>67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>SMW</strong></td>
<td>6.7</td>
<td>7.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>EMW</strong></td>
<td>20.9</td>
<td>21.6</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

E: estimated value in LHIN 4; A: actual value in Ontario; SMW: semi-urgent median wait time; EMW: elective median wait time; -: actual data is not available.

used in the calculation process are shown in Figure 5.6(b). The *age profile*, *RI profile*, *capacity*, and *supply* should be considered when calculating the value of *service utilization*, as the four exogenous LVs have significant direct effects on *service utilization* according to Figure 5.6(a). Two directly related LVs, *capacity* and *service utilization*, must be considered when calculating the values of *throughput* and *wait times*, as both LVs have significant direct effects. Other LVs, i.e., *age profile*, *RI profile*, *capacity*, and *supply* have significant indirect effects via *service utilization*, so their effects must also be included in the calculation.

The rate of change in the *age profile* in LHIN 4 is 0.073% from 2006 to 2010 and 0.093% from 2006 to 2011 [36], respectively. As detailed information about the *RI profile* in LHIN 4 is not available, we use the trends for Ontario as a whole instead. According to [36], the RI ratio is assumed to be 0.009 of the total population in Ontario since 2008. The estimated cardiac surgery utilization in LHIN 4 is shown in Table 5.2.

According to the estimation results shown in Table 5.2, the estimated throughput of cardiac surgery in LHIN 4 is higher than the actual average throughput in Ontario,
and the estimated elective wait times for cardiac surgery in LHIN 4 are almost equal to those in Ontario. This finding is in accordance with the observed performance pattern of cardiac surgery in LHIN 4, as the ratios of the throughput and elective wait times for cardiac surgery in LHIN 4 to those in Ontario are 1.42 and 1.02, respectively, according to the training data set. Our method is therefore able to accurately estimate cardiac surgery utilization with respect to demographic shifts.

(2) Simulation Results Based on the MSMQ-EC Queueing Model
To observe the dynamics of cardiac surgery performance in terms of the queue length and wait times in CS-ORs in 2010 and 2011, we show our simulation results based on the MSMQ-EC queueing model with the estimated service utilization. According to the CCN data, we assume that there are 50 patients waiting at the end of 2009. We initialize the queueing model with three servers (ORs), which can provide 1,400 cases annually, in accordance with the actual CS-OR operations in the HHSC. The simulation results for the queue length and average wait time from 2010 to 2011 are shown Figures 5.7 and 5.8.

![Graph showing queue length over time](image)

Figure 5.7: The estimated service utilization and resulting simulated queue lengths from 2010 to 2011.

As shown in Figure 5.7, the estimated queue lengths in 2010 and 2011 are different. The estimated queue length in 2010 does not increase beyond 60. The pattern does not hold in 2011, in which the simulated queue length shows a sharp increase. The longest queue length reaches 120 in 2011. Higher estimated service utilization in 2011 than in 2010 may account for the rising queue length in 2011.

According to Figure 5.8, the simulated average waiting time in 2011 is also longer than that in 2010. The average waiting time is around 7 days in 2010 and rises to almost 10 days at the end of 2011.
Figure 5.8: The estimated service utilization and resulting simulated average wait times from 2010 to 2011.

### 5.4 Summary

In this chapter, we focused on how to project the changes in healthcare service performance based on the underlying relationships between the demographic profiles, health service utilization, and service performance. We proposed an integrated projection method consisting of an SEM-based analysis, projection, and queueing model simulations, and tested the method on cardiac surgery services in Ontario, Canada. The results show that our proposed method can reveal the complex relationships between demographic profiles and healthcare service characteristics, enabling us to reasonably project the changes in service utilization and service performance with respect to demographic shifts. Our queueing models, which characterize certain operations within a healthcare service system, allow us to observe the dynamics of queue length and wait times in response to demographic shifts over time. This method will be helpful for a healthcare service system aiming to dynamically adjust its resources and management strategies, and thus maintain a stable service in terms of performance.
Chapter 6

Designing and Evaluating an Adaptive Strategy for Service Management

In the previous three chapters, we described how to investigate the complex effects of demand and supply factors on healthcare service performance in terms of wait times and how to estimate changes in service utilization and performance with respect to demographic shifts using the identified effects. In this chapter, we explore the design and evaluation of new strategies for improving service management behavior, which plays a significant role in service performance with respect to unpredictable patient arrivals. We use time block allocations in CS-ORs as a case study. We propose an adaptive OR time block allocation strategy using a self-organizing systems perspective that incorporates historical feedback information about ORs. We evaluate the performance of the proposed strategy using a queueing model derived from general perioperative practices and a discrete-event simulation study. Figure 6.1 summarizes the research focus of this chapter and how it fits into the larger context of understanding a healthcare service system, and how the study is conducted.
6.1 Introduction

The healthcare service system is a complex system [61, 64] consisting of numerous factors affecting the system’s performance/outcome, e.g., unpredictable patient arrivals (service utilization), service capacity, and service management; and interactions (coupling relationships) between the factors and the system’s performance/outcome. As one of the major cost areas in hospitals, the OR can also be viewed as a typical complex healthcare service system, consisting of a number of impact factors (e.g., unpredictable arrivals) and positive/negative relationships between those impact factors and OR performance (e.g., the positive effect of arrivals on the waiting time and queue length [29], and the positive effect...
of service capacity on arrivals [6]). Due to the nature of its complexity, researchers and healthcare administrators have realized that improving the healthcare service system from a self-organizing systems perspective has promise [61, 64]. In this chapter, we aim to improve the utilization of ORs by incorporating an adaptive OR time block allocation strategy proposed from a complex systems point of view.

Different resource management strategies have been proposed to improve the utilization of ORs with respect to different indicators, such as service throughput, average wait time, queue length, the number of bumped non-urgent surgeries, and the number of unused OR time blocks [208]. A common strategy is to improve the allocation of OR time blocks. Existing studies have attempted to improve OR time block allocation by (1) estimating surgery lengths more accurately (e.g., the time taken to perform a combined coronary artery bypass surgery is longer than a non-combined case), so that the size of time blocks can be assigned more reasonably [208]; (2) analyzing and controlling the factors that cause surgery delays [209], by reducing the delay of the first surgery to avoid the cancelation of following surgeries in a day, for example; or by (3) strategically arranging non-urgent surgeries, as they account for nearly 85% of all surgeries [210].

A challenging, basic question involved in OR time block allocation is how many OR time blocks should be reserved to cope with the unpredictable arrival of urgent patients. Reserving more time blocks than are actually needed may cause lower OR utilization, a longer waiting list, and longer wait times for non-urgent surgeries, whereas reserving insufficient time blocks may increase urgent patients’ risk, result in more bumped non-urgent surgeries, and prolong the wait times for those bumped cases.

Earlier studies have used mathematical methods (e.g., job shop scheduling models) to compute the optimal number of reserved urgent time blocks. The goal of these methods is to maximize OR time block utilization while minimizing the overtime or cancellation of surgeries [52].

In some Ontario hospitals, OR time blocks are distributed to surgeons based
on the allocations made in previous years and may only be reviewed two or three times a year. As this allocation strategy is relatively static, it may not cope well with actual patient arrivals. Patient arrivals are dynamic because of the number of impact factors involved, such as the weather and patients’ service utilization behavior [22]. Adaptively reserving time blocks in accordance with dynamic patient arrivals will therefore lead to better use of OR resources.

This chapter uses a complex systems perspective to propose an adaptive OR time block allocation strategy for coping with dynamic patient arrivals. We measure the effectiveness of our strategy using the number of bumped non-urgent surgeries, which are cancelled surgeries that are replaced by urgent surgeries, and the number of unused urgent time blocks, which are assigned to urgent surgeries in advance but not used. We believe that a more effective OR time block allocation strategy will improve OR utilization. We evaluate the performance of our strategy by building a multi-priority, multi-server, non-preemptive queueing model with an entrance control mechanism based on the general practice of CS-ORs in the HHSC\(^1\) in Ontario, and using it to carry out discrete-event simulations.

### 6.2 Designing an Adaptive OR Time Block Allocation Strategy

One way to allocate OR time blocks is to reserve a certain number of time blocks for urgent surgeries and assign the remaining time blocks to surgeons for non-urgent surgeries. The number of time blocks allocated for urgent surgeries and those allocated to surgeons are usually based on the allocation methods used in previous years [10]. Allocating ORs with such a relatively static strategy may not effectively use the OR resources as patient arrivals are unpredictable. We therefore propose an adaptive OR time block allocation strategy that incorporates the system’s feedback. As illustrated in Figure 6.2, the main idea behind our strategy

---

\(^1\)http://www.hhsc.ca/
is to periodically adjust the time blocks allocated for urgent surgeries based on the feedback information. In period $T$, the OR scheduler is fed the OR time block allocation, the numbers of bumped non-urgent surgeries, the unused urgent time blocks, and the dynamic arrivals in $T-1$.

![Diagram of OR scheduler with feedback mechanism](image)

**Figure 6.2:** The OR scheduler with a feedback mechanism.

The mechanism that is used by our adaptive strategy for adjusting the time blocks is shown in Figure 6.3. When the OR scheduler makes a decision on the allocation of time blocks for the coming period $T$, the information from the past period $T-1$ is fed back to the OR scheduler. If the number of bumped non-urgent surgeries is larger than a threshold $\theta_1$ in $T-1$, the scheduler increases the number of time blocks ($R_T$) for urgent surgeries by a step size of $\Delta p$ in $T$. If the number of unused urgent time blocks is larger than a threshold $\theta_2$, the scheduler decreases $R_T$ by a step size of $\Delta q$ in $T$. The thresholds $\theta_1$ and $\theta_2$ are defined by:

$$\theta_i = \frac{a_i \times \sigma \times \hat{T}}{\hat{T}} \quad (i \in \{1, 2\})$$  \hspace{1cm} (6.1)

where $a_i$ is a positive integer, $\hat{T}$ is a unit of time (one week here), and $\sigma$ is the
standard threshold in $\hat{T}$. $\theta_i$ presents the tolerance of the adaptive scheduler to the variations of the unused OR time blocks or bumped non-urgent time blocks in $T - 1$. Here, the tolerance means that if there are few bumped cases or unused time blocks, the time block allocation strategy in $T - 1$ works and does not need to be changed. As healthcare administrators’ tolerance to unused OR time blocks and that to bumped non-urgent time blocks may not be the same, we utilize $a_i$ to differentiate the two tolerances based on a standard threshold $\sigma$.

![Figure 6.3: The mechanism for updating the OR time blocks for urgent surgeries.](image)

### 6.3 Modeling OR Services

We examine the performance of our adaptive strategy by building a queueing model to simulate the queueing situations in ORs with respect to the arrival/service patterns, service discipline (e.g., priority based service discipline), and scheduling strategies. Our queueing model (shown in Figure 6.4) is based on the CS-ORs in the HHSC in 2004. We assume that there are two homogeneous (in terms of service rate) ORs with an average of two time blocks per day in each OR and that there are five working days per week. The 1,400 patient arrivals for cardiac surgeries each year are categorized into urgent ($U$), semi-urgent ($S$), and
elective (E) priority groups. According to the historical data from [24], the ratios of U, S, and E patients are 0.23, 0.6, and 0.17, respectively. Patient arrivals in winter are approximately one quarter greater than in other seasons, because of seasonal factors such as the weather. We assume that the arrival rate $\lambda_i$ of each priority group $i$ ($i \in \{U, S, E\}$) follows a Poisson distribution and that the service rate $\mu$ of each OR follows an exponential distribution, similar to previous work [96].

![Image of a multi-priority, multi-server, non-preemptive queueing model with an entrance control mechanism.](image)

Figure 6.4: A multi-priority, multi-server, non-preemptive queueing model with an entrance control mechanism.

According to the scheduling rule, U patients are settled immediately in an available OR as they have the highest priority. If none of the ORs are available, U patients bump the first prescheduled OR block for non-urgent surgery. In reality, a number of OR time blocks are reserved to cope with the U patients in a timely manner. In our model, we use $\delta_0$ to denote the initial number of time blocks reserved for urgent surgeries in a unit of time $\hat{T}$ ($\hat{T}=1$ week in this work). S and E patients are scheduled by surgeons following a priority-based service principle. New non-urgent (i.e., S and E) patients are assigned to a surgeon $j$ ($j \in [1, 6]$ in our case denotes one of the six surgeons) with a probability $p_{j,U}$ ($\bar{U}$ denotes
non-urgent patients) and then wait in the queue of surgeon $j$. In reality, surgeons normally perform non-urgent surgeries in time blocks allocated to them in advance. Therefore, we assume that a patient at the head of a queue $j$ will move to the OR with a probability $q_{j,t}$ at the next time step. It should be noted that $p_{j,t}$ and $q_{j,t}$ follow uniform distributions in our simulations.

The queueing model is implemented using the discrete-event simulation toolbox SimEvents integrated with MATLAB 2010. The parameters in the simulations are initialized with statistical data from the HHSC in 2004. The performance of the adaptive strategy and its sensitivity to the parameter settings are investigated in specific scenarios.

### 6.4 Simulation-Based Experiments

#### 6.4.1 Experimental Settings

The HHSC\(^2\) is one of the most comprehensive healthcare service systems in Ontario, Canada. Approximately 1,400 cardiac surgeries are performed each year in this hospital. In 2004, it had six specialized surgeons and two ORs. The reports by the Surgical Process Analysis and Improvement Expert Panel in Ontario\(^3\) and by the Office of the Auditor General of Ontario Hospitals [10] list the general rules for scheduling ORs in the HHSC. The CCN\(^4\) has released statistical data on the performance of the HHSC since 2004, including the number of surgeries completed in each month (throughput) and the number of patients waiting at the end of each month (queue length). A study in collaboration with the HHSC in 2004 [208] reported data on the average service time per surgery and the number of canceled surgeries. Table 6.1 summarizes the HHSC cardiac surgery data, which we use to initialize our simulations.

\(^2\)http://www.hamiltonhealthsciences.ca/


\(^4\)http://www.ccn.on.ca/ccn\_public/FormsHome/HomePage.aspx
Table 6.1: Cardiac surgery services in the HHSC in 2004

<table>
<thead>
<tr>
<th>Performance indicator</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queue length (at the end of a month)</td>
<td></td>
</tr>
<tr>
<td>Quarter 1</td>
<td>156</td>
</tr>
<tr>
<td>Quarter 2</td>
<td>159</td>
</tr>
<tr>
<td>Quarter 3</td>
<td>149</td>
</tr>
<tr>
<td>Quarter 4</td>
<td>147</td>
</tr>
<tr>
<td>Cancellations</td>
<td></td>
</tr>
<tr>
<td>Bumped non-urgent surgeries</td>
<td>77</td>
</tr>
<tr>
<td>Service time</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>4.6 hours</td>
</tr>
</tbody>
</table>

6.4.2 Experimental Results

We first investigate the effect of our adaptive strategy on the queue length over a 3-year period (i.e., 156 weeks, where 13 weeks represents one quarter in the simulation and the third quarter each year corresponds to the winter season) using the queueing model. As shown in Figure 6.5, the average queue length with the adaptive strategy is slightly shorter than without the adaptive strategy. The results are obtained from a single simulation run with the patient arrival patterns described in the preceding section and the parameter settings as given in the figure caption. The queue lengths at the initial time step ($T = 0$) in the simulation are both set to 156, which is the number of patients who were waiting for cardiac surgeries at the end of 2003 in the HHSC. The fluctuations in the queue lengths, especially the increasing periods in the weeks 27-39, 79-91, and 131-143, correspond to the increased patient arrivals in the winter compared to the other seasons.

The existing OR time block allocation strategy depends heavily on the allocation methods used in previous years. A hidden assumption behind this strategy is that patient arrivals do not vary much within a year. This assumption
may not hold in reality as patient arrivals are dynamic in response to the complex environment (e.g., the weather) and patients' personal behavior. The static time block allocation strategy may therefore result in a number of bumped non-urgent surgeries when there are more urgent arrivals, or it may lead to under-utilization of OR time blocks due to fewer urgent arrivals. The strategy proposed in this chapter adaptively adjusts the OR time block based on historical information, for example, the utilization of ORs in the previous week/month/quarter and the number of urgent/non-urgent arrivals.

### 6.5 Discussion

We conduct several additional simulation experiments with different parameter settings (i.e., the initial reserved blocks for urgent surgeries in a week, the adjustment time interval, threshold, and step size) to confirm the observations made. Figure 6.6 shows that without the adaptive strategy, the number of bumped non-urgent surgeries drops and the number of unused time blocks increases when
\( \delta_0 \) increases. This finding suggests that the utilization of ORs without the adaptive strategy is more sensitive to the number of time blocks allocated to urgent surgeries, whereas the adaptive strategy is robust to the initial number of OR blocks for urgent surgeries. The figure also shows that the OR can maintain a trade-off between the number of bumped non-urgent surgeries and the number of unused urgent time blocks with the adaptive strategy. This finding implies that hospitals can adapt to dynamically changing patient arrivals with our adaptive strategy and hence can improve their OR utilization. Furthermore, Figure 6.6 reveals that OR utilization is improved when \( \delta_0 \) is set to 5-8, in the scenario defined by the parameter settings in the caption of Figure 6.6.

The time interval \( T \) for allocating OR time blocks (e.g., once per week/month/quarter) is another key parameter in the adaptive strategy. Figure 6.7 shows the effects of different updating time intervals (\( T \)) on OR utilization in terms of the trade-offs between the numbers of bumped non-urgent surgeries and unused urgent time blocks. We measure the trade-off with BNS/UUB. OR utilization improves as the measure approaches 1 (represented by the dotted line in Figure 6.7). As Figure 6.7 shows, updating the OR time block once every 4-8 weeks will both reduce the number of bumped non-urgent surgeries and balance the number of unused urgent blocks. In the scenario in this figure, updating the OR time block once every 4 weeks will result in the best OR utilization.

The adjustment thresholds (\( \theta_1 \) and \( \theta_2 \)) and the step sizes (\( \Delta p \) and \( \Delta q \)) may also affect the performance of the adaptive strategy. According to Figure 6.8, larger adjustment thresholds (i.e., larger \( \sigma \)) result in a larger number of unused urgent time blocks and a smaller number of bumped non-urgent surgeries. This is reasonable, as seven time blocks are initially reserved for urgent surgeries, which almost satisfies the average number of urgent arrivals in a week according to the patient arrival patterns. Intuitively, larger thresholds make ORs less likely to increase or decrease the time blocks for urgent surgeries, and vice versa. In such cases, the adaptive
The Utilization of ORs (No.)

Figure 6.6: OR utilization with respect to different initial urgent OR time blocks (AS: adaptive strategy; BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; $\delta_0 = 7$ per week, $T = 4$ weeks, $\theta_1 = 2 * T$, $\theta_2 = 1 * T$, $\Delta p =, \Delta q = 1$).

strategy will be less flexible and hence can lead to a worse OR utilization.

Figure 6.7: The trade-offs (i.e., BNS/UUB) of the adaptive strategy with respect to different $T$ (BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; $\delta_0 = 7$ per week, $\theta_1 = 2 * T$, $\theta_2 = 1 * T$, $\Delta p =, \Delta q = 1$).

Figure 6.9 shows that smaller step sizes (e.g., $\Delta p$ and $\Delta q$ are set to 1 or 2) can guarantee better OR utilization in the given specific scenario. Intuitively, a larger
Figure 6.8: OR utilization with respect to adjustment thresholds (BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; δ₀ = 7 per week, T = 4 weeks, θ₁ = 2 * σ * T, θ₂ = σ * T, ∆p =, ∆q = 1).

A step size will lead to the number of time blocks for urgent surgeries increasing or decreasing by a larger amount at a time, and thus will result in a larger number of unused urgent time blocks or bumped non-urgent surgeries in the next time step.

We can fine-tune the parameter settings of the adaptive strategy using the above results. Figure 6.10 presents a comparison of the queue lengths generated by the original adaptive strategy (defined by Setting I, which is the same as the setting configuration of the adaptive strategy in Figure 6.5), the fine-tuned adaptive strategy (defined by Setting II), and by the allocation schedule without the adaptive strategy. The fine-tuned adaptive strategy has a shorter queue length than the original adaptive strategy most of the time.
The Utilization of ORs (No.)

Step Size ($\Delta p, \Delta q$)

Figure 6.9: OR utilization with respect to step sizes (BNS: bumped non-urgent surgeries; UUB: unused urgent time blocks; $\delta_0 = 7$ per week, $T = 4$ weeks, $\theta_1 = 2T$, $\theta_2 = T$).

Figure 6.10: Queue lengths generated by different block allocation strategies (Setting I: $\delta_0 = 7$ per week, $T = 4$ weeks, $\theta_1 = 2T$, $\theta_2 = T$, $\Delta p =, \Delta q = 1$; Setting II: $\delta_0 = 5$ per week, $T = 4$ weeks, $\theta_1 = 2T$, $\theta_2 = T$, $\Delta p =, \Delta q = 1$).

6.6 Summary

In this chapter, we proposed an adaptive strategy for allocating OR time blocks based on a feedback mechanism, as a case study for how to improve healthcare
service system management for better service utilization and performance. We evaluated the effectiveness of the adaptive strategy in improving the utilization of service resources using a specific multi-priority, multi-server, non-preemptive queueing model with an entrance control mechanism based on the general perioperative process of the CS-ORs in the HHSC. By applying the adaptive strategy to this queueing model, we showed that our adaptive strategy is able to efficiently regulate OR time block reservations in response to dynamical changes in patient arrivals. The adaptive strategy could maintain a better trade-off between the number of bumped non-urgent surgeries and the number of unused urgent OR time blocks, leading to shorter waiting lists and wait times. The proposed adaptive strategy suggests that frequently adjusting the OR time block allocation (i.e., once per month) is helpful for improving OR utilization.
Chapter 7

Characterizing the Tempo-Spatial Patterns in Patient Arrivals and Wait Times

In the previous four chapters, we studied how to identify the complex effects that demand or supply factors exert on wait times, how to estimate wait time dynamics with respect to changes in certain demand factors, and how to design service management strategies to improve wait times. We have not yet explored how to model and characterize emergent tempo-spatial patterns at a systems level by taking into account the underlying entities’ behavior (e.g., patient hospital selection behavior) with respect to various impact factors (e.g., the distance between homes and services, hospital resourcefulness, and historical wait time information). In this chapter, we use a behavior-based autonomy-oriented modeling method to model and simulate service utilization in cardiac surgery services and use the resulting model to characterize self-regulating patient arrivals and wait times. By experimenting with the Autonomy-Oriented Computing (AOC)-based cardiac surgery service model (AOC-CSS model), we reveal the working mechanisms that explain how the tempo-spatial patterns in patient arrivals and wait times at a systems level emerge from individual patients’ hospital
selection behavior and their relationships with hospital wait times. Figure 7.1 summarizes the research focus of this chapter within the larger context of understanding a healthcare service system and the corresponding research steps.

![Diagram](a)

Figure 7.1: A schematic diagram illustrating the use of behavior-based autonomy-oriented modeling to characterize the emergent tempo-spatial patterns in patient arrivals and wait times in a healthcare service system. (a) The research focus of this chapter (highlighted in red) is part of the larger context of understanding a healthcare service system. (b) Modeling, simulating, and analyzing the behavior of a healthcare service system.

### 7.1 Introduction

A healthcare service system, such as the cardiac care system schematically illustrated in Figure 7.2, is well recognized as a complex system [61, 64]. Some interesting self-organizing tempo-spatial patterns in healthcare service utilization,
such as the power-law distribution of variations in specialists’ waiting lists (i.e., the variations in the mean time that patients spend on specialists’ waiting lists) [6], have been reported. However, it is still unclear what kinds of individual behavior (e.g., patient hospital selection behavior and hospital service adjustment behavior) and which underlying factors (e.g., distance between homes and services, hospital resourcefulness in terms of physician supply, and service performance, represented by wait times) account for these emergent tempo-spatial patterns and how they accomplish these patterns.

We use a behavior-based autonomy-oriented modeling method to understand certain tempo-spatial patterns relating patient arrivals and wait times from a complex systems self-organizing perspective, focusing on cardiac surgery services. To model the real-world cardiac surgery system in Ontario, Canada, the essential issues of scope, coupling relationships, and heterogeneity must be addressed.

- **Scope**: What factors, entities, processes, and hierarchical levels (e.g., services at a hospital or regional level) are relevant to the tempo-spatial patterns, and hence should be investigated and modeled?

- **Coupling relationships and/or interactions**: What are the relationships between the impact factors and variables? Identifying the local feedback loop(s) will be crucial for understanding global-level self-organized regularities.

- **Heterogeneity**: Patient behavior when choosing a hospital may be heterogeneous, due to the differences in personal profiles, socioeconomic backgrounds, and service distributions in and around their residence areas. Hospitals may also be heterogeneous in delivering healthcare services, because of variations in equipped resources, management strategies, and dynamically changing patient arrivals. Thus, capturing the heterogeneity of patients and hospitals is essential in modeling and simulating a real-world complex healthcare system.
We use a behavior-based autonomy-oriented modeling method [57] to construct an AOC-CSS model. In modeling the real-world cardiac care system in Ontario, Canada, we consider multiple factors affecting patient arrivals (shown in Figure 7.2), such as the weather, the demographics of cities and towns in Ontario, the geographic accessibility of cardiac surgery services, the resourcefulness of physicians in a hospital, hospital performance in terms of wait times, and patients’ hospital selection behavior.

Following a behavior-based autonomy-oriented modeling method, we introduce the tempo-spatial patterns in patient arrivals and wait times, which are observed from the aggregated data on cardiac surgery services. We identify the key entities, major factors, and local feedback loops that should be modeled. We present the detailed formulation of the developed AOC-CSS model. We use the model in simulation-based studies and present results characterizing the regularities in patient arrivals and hospital wait times. We discuss the underlying mechanism that is revealed by the validated AOC-CSS model and a sensitivity analysis on the key parameters that influence the emergence of self-organized patterns.
Figure 7.2: The complex cardiac surgery system in Ontario, Canada. The illustrated tempo-spatial patterns are observed from the secondary data on cardiac surgery services between January 2005 and December 2006. LHIN: Local Health Integration Network; H: hospital.
7.2 Empirical Tempo-Spatial Patterns in Cardiac Surgery Services

Prior studies have empirically identified self-organized regularities in healthcare systems. For instance, Smethurst and Williams found that the monthly absolute variations in the time that patients spend on specialists’ waiting lists (calculated as the change in the mean wait time \( w \) at time steps \( t \) and \( t - 1 \) \( (w_t - w_{t-1})/w_t \) followed a power-law distribution [6] and concluded that hospital waiting lists were self-regulating. We aim to discover the corresponding patterns in the cardiac surgery services from empirical data, focusing on three research questions.

1. What are the statistical distributions of the variations in the number of patient arrivals and wait times?

2. What are the spatial patterns of patient flows? Are there any underlying patterns that may be observed from the spatial distribution of patient flows?

3. What are the temporal patterns in patient arrivals and wait times?

We once again focus on the cardiac care system in the province of Ontario, Canada, specifically the Ontario LHINs\(^1\). Each LHIN is a geographic-location-based, sub-provincial administrative unit responsible for determining the healthcare service needs and priorities for its corresponding area. In Ontario, there are 14 LHINs, differing in their administrative areas, geographic sizes (shown in Table 3.1), and geodemographic profiles (shown in Table 3.2). The cardiac care services which we are interested in, for example the cardiac surgery services, are unevenly distributed across the 14 LHINs. Figure 7.3 illustrates the distribution of the 11 hospitals that provide cardiac surgery services across the 14 LHINs. In this section, we introduce the collection and preprocessing of the aggregated data for this investigation, and the specific self-organized patterns in patient arrivals and wait times that are discovered from the empirical data.

\(^1\)http://www.lhins.on.ca/home.aspx
Figure 7.3: The distribution of hospitals that provide cardiac surgery services across the LHINs in Ontario, Canada (adapted from the Ontario LHINs map given at http://www.lhins.on.ca). Red dots denote hospital locations and the numbers correspond to the LHIN IDs, following Table 3.3.

7.2.1 Aggregated Data

The data used for investigating self-organized self-regularities in a real-world cardiac surgery system were obtained from the CCN\(^2\). As a support system to cardiac care in Ontario, the CCN is a network of 18 member hospitals providing cardiac care services, of which 11 hospitals provide cardiac surgery services (illustrated in Figure 7.3). The CCN has published monthly wait time information for cardiac surgery services in member hospitals across Ontario between January 2005 to December 2006. We accessed the data in February 2011. As shown in Table 4.2, the reported CCN data include the number of completed cases in a month (i.e., throughput), the average number of patients waiting at the end of a month (i.e., queue length), and the monthly median wait time for urgent,

\(^2\)http://www.ccn.on.ca/
semi-urgent, and elective patients. We use the median wait time for elective patients (referred to hereafter as the median wait time) to measure wait times, which may represent the changes in the overall wait times for cardiac surgery services better than the 90% wait times.

Based on the CCN data, we are able to calculate the average monthly number of patient arrivals at each hospital by:

\[
Arrival_i^t = Throughput_i^t + NoWait_i^t - NoWait_i^{t-1},
\]

(7.1)

where \( Arrival_i^t \) is the average monthly number of arrivals in quarter \( t \) of unit \( i \), \( Throughput_i^t \) is the average monthly number of patients who have received treatment in quarter \( t \) of unit \( i \), and \( NoWait_i^t \) is the average number of patients waiting at the end of a month in quarter \( t \) of unit \( i \).

We use the above-described data for cardiac surgery services over two years, from January 2005 to December 2006, to discover the self-organized patterns in patient arrivals and wait times.

### 7.2.2 Statistical Regularities

Following the work of Smethurst and Williams [6, 70], we first investigate the statistical distributions of the variations in patient arrivals and wait times. From the CCN data, the month-to-month variations in patient arrivals and wait times are calculated by:

\[
v_{t+1} = \frac{x_{t+1} - x_{\min}}{x_{\max} - x_{\min}} - \frac{x_t - x_{\min}}{x_{\max} - x_{\min}},
\]

(7.2)

where \( v_{t+1} \) denotes the variation in patient arrivals or wait times at time \( t + 1 \), \( x_t \) denotes the number of patient arrivals or the wait times at time \( t \), \( x_{\min} \) and \( x_{\max} \) are the minimum and the maximum values of patient arrivals or wait times over the two-year period, respectively. In this work, each time step \( t \) corresponds to a month.
The absolute month-to-month variations in patient arrivals or wait times, $v'_t$, are then calculated by:

$$v'_t = |v_t|. \quad (7.3)$$

Two types of self-organized regularities are identified from the (absolute) variations in patient arrivals and wait times, as shown in Figures 7.4 and 7.5. As shown in Figure 7.4, the monthly variations in patient arrivals against the percentage of variation occurrences follow a normal distribution with a mean value of 0.004 and a standard deviation of 0.226. The normality of the distribution passes the Kolmogorov-Smirnov test [211, p.392-394].

![Figure 7.4: The statistical distribution of variations in patient arrivals for cardiac surgery services in Ontario, Canada, between January 2005 and December 2006. The normality of the distribution passes the Kolmogorov-Smirnov test. Standard deviation $SD = 0.226$.](image)

As shown in Figure 7.5, the monthly absolute variations in the median wait time follow a power-law distribution with a power of $-1.36$ and a standard deviation of 0.28 (linear fitness: $p < 0.001$). The fitness of the power-law distribution is
tested using the method proposed by Clauset et al. [212] (power-law test: $p < 0.1$).

The median wait time for cardiac surgery services therefore exhibits a statistical regularity in its month-to-month variations, suggesting that the cardiac care system is, to a degree, able to self-organize [213] its wait times.

Figure 7.5: The statistical distribution of the absolute variations in median wait times for cardiac surgery services in Ontario, Canada, between January 2005 and December 2006. The distribution follows a power law with a power of $-1.36$. (power-law test based on Clauset’s method [215]: $p < 0.1$; linear fitness (red line): $p < 0.001$; standard deviation $SD = 0.28$).

### 7.2.3 Spatial Patterns

(1) **Patient Flow Distributions**

Figure 7.6 shows the distribution of the number of patients residing in each LHIN against the LHINs where they receive cardiac surgery services between 2007 and 2008 in Ontario, Canada [136]. The distribution can be regarded as the spatial pattern of patient flows, which represents the aggregated effects of patients’
hospital selection behavior. We find approximately the same spatial pattern using the reported statistical data from 2007 to 2011 [35, 136, 214, 215, 216]. The percentage of cardiac surgery patients operated in an LHIN with respect to their LHIN residence varies within 5% year to year over the four years, with a maximum value of approximately 10%.

![Diagram of cardiac surgery services location by LHIN](image)

**Figure 7.6:** The distribution of cardiac surgery patients with respect to their LHIN residence between 2007 and 2008 in Ontario, Canada.

**(2) Patient-Attraction and Patient-Distribution Degrees for LHINs**

The spatial pattern of patient flows shown in Figure 7.6 may represent additional information about the probability that an LHIN attracts patients residing in other LHINs, called the patient-attraction degree, or the probability that patients living in a specific LHIN travel to other LHINs for services, called the patient-distribution degree. A higher patient-attraction degree indicates that the LHIN is linked by more patients from other LHINs and reveals how heavily patients from other LHINs can affect the arrivals for the hospital(s) in a specific LHIN. A higher patient-distribution degree indicates that patients living in that LHIN are more likely to disperse to other LHINs to receive cardiac surgery services. This in turn reveals the extent to which patients in a specific LHIN may influence arrivals at hospitals in other
LHINs. In this section, we introduce how to reveal this underlying information and the corresponding patterns.

(i) Calculation Method

We use the idea behind the hyperlink-induced topic search (HITS) algorithm designed by Kleinberg [217] to calculate the patient-attraction degree and the patient-distribution degree for each LHIN from the patient flow distribution. The HITS algorithm is a linkage structure-based analysis algorithm. It characterizes to what extent a web page is an “authority,” by estimating the in-degree of a page, or to what extent it is a “hub,” by estimating the out-degree of a page, based on the relationships between a set of related web pages.

Our proposed method is similar to the HITS algorithm. Based on the distribution of patient flows, LHINs can form a network structure where a directed link between two LHINs indicates that there are patients coming from the LHIN the link points away from, to the LHIN the link points toward. Thus, an LHIN can be regarded as analogous to a web page, its patient-attraction degree as analogous to the “authority” value of a web page, and its patient-distribution degree as analogous to the “hub” value of a web page. We can therefore estimate the values of the patient-attraction degree and patient-distribution degree for each LHIN via the eigenvectors of the matrices associated with the distribution of patient flows, as proposed by Kleinberg [217].

Given a patient-flow matrix $F$, each entry $(i, j)$ of $F$ is the percentage of patients residing in LHIN $i$ and receiving cardiac surgery services in LHIN $j$. Based on Kleinberg’s theorem [217, p.11], the patient-attraction degree vector (i.e., the authority weight vector in the HITS algorithm) is the principle eigenvector of $F^T F$ and the patient-distribution degree vector (i.e., the hub weight vector in the HITS algorithm) is the principle eigenvector of $FF^T$. We use the patient flow distribution between 2007 and 2008, using data obtained from [136], to calculate the patient-attraction degree and the patient-distribution degree for each LHIN.

(ii) Patient-Attraction and Patient-Distribution Degrees
We calculate the patient-distribution degree and the patient-attraction degree for each LHIN using the method described about, shown in Figure 7.7. From the patient-distribution degrees shown in Figure 7.7, we observe that LHINs can be roughly classified into two groups, a high-patient-distribution group and a low-patient-distribution group. LHINs in the high-distributed group (LHINs 5-9 and 12) have obviously larger distributed degrees than those in the low-distributed group (LHINs 1-4, 10, 11, 13, and 14). These two distributed groups may be related to the geographic accessibility to services (service accessibility) for each LHIN. The LHINs in the high-distributed group are more accessible to cardiac surgery services than the LHINs in the low-distributed group, according to the service accessibilities presented in Table 3.2.

![Figure 7.7](image)

Figure 7.7: The distribution of LHINs’ patient-attraction and patient-distribution degrees.

From the patient-attraction degree for each LHIN shown in Figure 7.7, we again observe that the LHINs can be roughly classified into two groups, a high-patient-attraction group and a low-patient-attraction group. The LHINs in the high-patient-attraction group (LHINs 6-8) have more patients traveling from other LHINs than the LHINs in the low-patient-attraction group (LHINs 1-5 and 9-14). The formation of the two groups may be driven by the geographic locations
and reputations (e.g., the number of physicians and wait times) of the hospitals in each LHIN. For instance, LHIN 7 has the highest patient-attraction degree and is the only LHIN with three hospitals, which all have sufficient personnel and facilities. LHIN 6 and 8’s relatively higher patient-attraction degrees could be because these LHINs have hospitals providing cardiac surgeries and they have one or more neighboring LHINs that lack cardiac surgery services (e.g., LHIN 5, 9, and 12), as shown in Figure 7.3.

### 7.2.4 Temporal Patterns

As patient arrivals and wait times change dynamically, there may be temporal regularities, in addition to the statistical and spatial patterns already identified. We aim to identify any existing temporal patterns, focusing on two specific research questions.

1. What are the changing trends in patient arrivals and wait times? Are there any patterns in the temporal variations in patient arrivals and wait times, such as monthly or seasonal patterns?

2. As historical information about the wait times in each hospital is expected to affect the subsequent patient arrivals, does the cardiac care system exhibit patterns that reveal the potential interactions between patient arrivals and wait times?

Figures 7.8 and 7.9 show the monthly variations in patient arrivals and wait times, respectively, in cardiac surgery services in Ontario from 2005 to 2006. Although patient arrivals for cardiac surgery services fluctuate from month to month in both Ontario (shown in Figure 7.8(a)) and in each hospital (shown in Figure 7.8(b)), there is a seasonal pattern. There are relatively smaller numbers of patient arrivals for cardiac surgery services in the warm season (shown as the shadowed areas in Figure 7.8(a)), which runs from the fifth month (May) to the eighth month (August), than in the colder months.
Figure 7.8: The changes in patient arrivals for cardiac surgery services in (a) Ontario and (b) five hospitals that provide cardiac surgery services.
Figure 7.9: The changes in the median wait times for cardiac surgery services in (a) Ontario and (b) five hospitals that provide cardiac surgery services.
We observe a similar seasonal pattern in the dynamically changing wait times in Figure 7.9. The median wait times in Ontario consistently decrease from the sixth month (June) to the tenth month (October) each year, shadowed in Figure 7.9. The lower wait times may be due to the lower patient numbers in cardiac surgery services in the warm season, as illustrated in Figure 7.8(a).

### 7.3 AOC-CSS Modeling

#### 7.3.1 Identifying Key Elements in Modeling

1. **Entities**

   In Ontario, each location (e.g., a city or a town) has a certain number of patients that require cardiac surgery services. When these patients are recommended to have cardiac surgery by their GPs or specialists, they will choose a specific hospital to receive the required services from [218]. In most cases, patients make their decisions with their GPs, as 93% of Ontario’s population are registered with a GP [219] and most of the patients will follow a GP’s recommendations [34, 220]. Patient hospital selection behavior therefore represents the consequence of a patient-GP mutual decision. After patients make a decision on hospital selection, they visit the selected hospital and wait to receive the treatment [218]. Finally, the patient leaves the hospital after finishing the treatment. We can therefore identify three entities in the cardiac care system, the *patient*, the *GP*, and the *hospital*.

2. **Major Impact Factors**

   Dynamically changing patient arrivals and wait times may be directly or indirectly affected by various factors. These factors, as illustrated in Figure 7.2, can be divided into two categories.

   - Factors affecting the patient population: Factors such as environment (e.g., weather), demographics (e.g., population size and age), and socioeconomics
(e.g., education), may affect the number of patients who have cardiovascular disease. Thus, we consider these factors when initializing the parameter of generating patient population for each city or town in simulations.

- Factors affecting the dynamics of patient arrivals to hospitals: Factors such as the geographic distance between homes to a hospital, hospital reputation (e.g., hospital resourcefulness), hospital performance (e.g., wait times), and decision making style may affect the patients’ hospital selection behavior, and thus result in the variations of patient arrivals to each hospital. We therefore consider these factors when designing behavioral rules for patients and hospitals.

In our modeling, based on the literature and our SEM-based studies (please refer to Chapters 3 and 4 for these studies), the factors that affect the patient population and thus should be considered in the simulation initialization are summarized below.

- Geodemographic profile of a location: As we investigated in Chapter 3, cities and towns with distinct geodemographic factors have different patient arrivals for cardiac surgery services. We therefore consider the differences in the geodemographic profiles of locations, which are represented by the patient arrival rate in the modeling.

- Seasonal weather: Seasonal weather is an important contributing factor for the outbreak of many diseases, including cardiac diseases [22], and therefore influences patient arrivals and wait times in cardiac care services. For instance, as shown in Figure 7.6, the patient arrival rate in the warm season (from May to October) in Ontario is approximately 15% lower than that in the cold season (from January to April and from November to December), according to the reported CCN data [136]. We therefore consider the factor of seasonal weather in our modeling, which is represented in different arrival rates in warm and cold seasons.
The identified major factors that influence the patient behavior in selecting hospitals and the hospital behavior in delivering services and thus are considered in our modeling are summarized below.

- **Geographic distance**: As revealed by our SEM-based analysis (please refer to Chapter 3 for the study) and the literature, the geographic distance between homes and a hospital is negatively associated with the probability that patients and GPs select a hospital [12, 221], because patients are more likely to visit hospitals close to their homes. Thus, we take into account the geographic distance in modeling patients’ hospital selection behavior.

- **Hospital resourcefulness**: The resourcefulness of a hospital, represented by the number of physicians [222] in this study, is positively correlated with the probability that patients and GPs select a specific hospital [222, 223, 224] because more hospital resources may attract more patient arrivals [6]. We therefore consider this factor when designing behavioral rules for patients’ hospital selection behavior.

- **Hospital performance in terms of wait times**: Wait times for receiving the required cardiac care services are a major concern for patients [34] and GPs [12, 225], who are usually in favor of hospitals with short wait times [12, 34, 225]. We therefore take this factor into consideration when designing behavioral rules for patients.

(3) **Local Feedback Loops**

The impact factors of wait times may have complex relationships, coupled interactions, and/or feedback loops [53]. These interactions, especially the local feedback loops, may result in nonlinear phenomena (e.g., self-regulating patient arrivals and wait times) in the complex cardiac care system.

We identify two local feedback loops between the impact factors, shown in Figure 7.10. The first negative feedback loop (namely AW-loop) exists between the
factors of patient arrivals and wait times, due to the patient-GP mutual decisions on hospital selection. For instance, long wait times in a hospital may weaken the probability of patients and GPs selecting that hospital, which will in turn decrease the number of patient arrivals and result in a decrease in wait times.

Figure 7.10: The effects of impact factors on patient-GP mutual decisions on hospital selection and the local feedback loops. “+”: positive relationship between two factors; “−”: negative relationship between two factors.

As shown in Figure 7.10, the factors of patient arrivals, hospital service rate, and wait times form a positive feedback loop (named ASW-loop) due to hospitals’ service adjustment behavior. If there are more patient arrivals at a hospital, that hospital will increase its service rate. Wait times will therefore decrease, which will in turn result in a larger number of patient arrivals.

7.3.2 Modeling Environment

In this study, the geographic relationship/structure between patients’ locations and hospitals is conceptualized as a bipartite location-hospital network $CH$, defined below.

**Definition 7.1 (Location-hospital network):** A location-hospital network can be described as a bipartite network $CH = (C, H, F, I)$. The location node set $C(N) =$
\{c_i\} (i \in [1, N]) denotes N cities, towns, or concerned sub-regions as patients’ locations. The hospital node set \(H(M) = \{h_j\} (j \in [1, M])\) represents M hospitals that provide specific healthcare services, \(H \cap C = \emptyset\). The adjacent matrix \(F = \{f_{ij}\}_{N \times M} (f_{ij} \in [0, 1], \sum_{j \in [1, M]} f_{ij} = 1)\) represents whether or not there are patient flows between each pair of city-hospital nodes. \(I = \{inf_{ij}\}_{N \times M}\) represents the static or dynamic information between each pair of city-hospital nodes.

Here, each location node \(c_i (\forall c_i \in C)\) represents a city or town with a population of more than 40,000 in 2006 in Ontario, Canada, according to census data. Each hospital node \(h_j (\forall h_j \in H)\) denotes a hospital that provides cardiac surgery services in Ontario, Canada. The location-hospital information is defined as \(I = \{inf_{ij}(t)|i \in [1, N], j \in [1, M]\}\) = \{\(d_{ij}\)|i \in [1, N], j \in [1, M]\}, where \(d_{ij}\) represents the distance from a city or town \(c_i (\forall c_i \in C)\) to a hospital \(h_j (\forall h_j \in H)\). Following Chapter 3, the distance \(d_{ij}\) is represented by the driving time between a city or town and a hospital. The driving time is again estimated using the “Get directions” function in Google Maps \(^3\).

Based on the location-hospital network \(CH\), the environment \(E\) in the AOC-CSS model records the released information about hospitals. We formally define the environment \(E\) as described below.

**Definition 7.2 (Environment):** The environment \(E\) for the AOC-CSS model is represented by a bipartite network, as defined in Definition 7.1. \(E\) maintains information that can be accessed by patients and GPs. We define the environment \(E\) as a tuple \(<D, R, W>\), where the elements are defined as follows:

- **\(D\): Distance information** \(D = \{d_{ij}|i \in [1, N], j \in [1, M]\}\). Each \(d_{ij}\) records the driving time between a city/town \(c_i (\forall c_i \in C)\) and a hospital \(h_j (\forall h_j \in H)\).

- **\(R\): Hospital resourcefulness information** \(R = \{r_j|j \in [1, M]\}\), where \(r_j\) records the number of physicians in \(h_j (\forall h_j \in H)\).

\(^3https://maps.google.com/\)
• \( W \): Wait time information \( W = \{w_{j,\tau} | j \in [1, M]\} \). Each \( w_{j,\tau} \) records the wait time information for a hospital \( h_j (\forall h_j \in H) \) at time round \( \tau \). Here, a unit time round \( \tau \) to review hospital operations (e.g., one month or one quarter) includes \( T \) number of unit time steps \( t \) (a unit of time to record the hospital operational information, e.g., one day), i.e., \( \tau = T \times t \). In this paper, \( w_{j,\tau} \) records the median wait time of \( h_j \) over the past time round \( \tau - 1 \).

### 7.3.3 Modeling Entities

**1. Patient**

As reported in [34], a large number of patients may not have access to wait time information and thus they may not consider wait times when they select a hospital. Patients can therefore be categorized as *wait time-sensitive* or *wait time-insensitive*, according to their decision making styles. *Wait time-sensitive* patients consider all of the acquired information about the hospitals (i.e., distance, hospital resourcefulness, and wait times). *Wait time-insensitive* patients do not take into account wait time information when they select hospitals. A patient entity is defined as described below.

**Definition 7.3 (Patient entity)**: A patient entity, \( \text{patient} \), maintains a record: \(< \text{patientID}, \text{cityID}, P_r, \text{rule}, \text{hospitalID}, \text{type}, \text{joinTime}, \text{endTime}, \bar{w} >\) , where the elements are defined as follows:

- **patientID**: This records the unique identity represented by a constant for a patient.

- **cityID**: This denotes the unique identity for the city/town that a patient comes from.

- **\( P_r \)**: This denotes the probability of a patient considering the factor of wait times when selecting a hospital. Accordingly, the probability of a patient who
does not take into account the factor of wait times when choosing a hospital is 
$1 - P_r$.

• rule: This indicates how a patient chooses a hospital along with the GP.

• hospitalID: This indicates the unique identity for the hospital that a patient arrives at.

• type: This represents the urgency of a patient entity to the cardiac surgery service according to the severity of illness, $\forall k \in [1, K]$ ($K \geq 1$). In this study, there are two urgent types: urgent patients and non-urgent patients.

• joinTime: This denotes the time step that a patient joins in the queue of a hospital.

• endTime: This indicates the time step that a patient has been served in a hospital.

• $\tilde{w}$: This records the wait time information of a patient, $\tilde{w} = \text{endTime} - \text{joinTime}$.

(2) GP

In the AOC-CSS model, patients come to a hospital that is selected by patient-GP mutual decisions and the released information in the environment $E$. As most cardiac surgery patients are referred by GPs, we define entities $GP[N]$ to record and represent patient-GP mutual decisions on hospital selection, as described below.

**Definition 7.4 (GP entity):** $GP[N]$ records the information about patients who live in specific locations and receive cardiac surgery services. Each entity $GP_i$ ($i \in [1, N]$) maintains a record: $<\text{cityID}, A_k(t) >$, where the elements are defined as follows:

• cityID: This represents the unique identity of a location.
• $A_k(t)$: This denotes the patient flow information for urgent type $k$ ($k \in K$) patients, $A_k(t) = \{\hat{a}_{k,j}(t)\}$. Each $\hat{a}_{k,j}(t)$ records the number of type $k$ ($k \in K$) patients to hospital $h_j$ ($h_j \in H$) at time step $t$.

(3) Hospital

We model the operations of a hospital entity based on queuing theory. As CS-ORs in a hospital are, to a certain extent, homogeneous, it is reasonable to regard a hospital $j$ as one server (i.e., one OR) with a service rate $\mu_j$, and thus assume that each hospital is an M/M/1 queuing model [226]. A hospital entity is defined as described below.

**Definition 7.5 (Hospital entity):** Hospital[M] records the information on all of the hospitals. Each hospital entity $h_j$ ($\forall h_j \in H$) maintains a record $<\text{hospital ID}, \text{city ID}, \tilde{A}_k(t), \mu(t), \text{rule}, w(\tau), \text{queue}>$, where the elements are defined as follows:

• **hospital ID:** This represents the unique identity for a hospital.

• **city ID:** This indicates the unique identity for the city/town in which a hospital is located.

• $\tilde{A}(t)_k$: This records the patient arrival information for type $k$ ($k \in K$) patients, $\tilde{A}(t)_k = \{\tilde{a}_{i,k}(t)\}$. Each $\tilde{a}_{i,k}(t)$ records the number of type $k$ ($k \in K$) patients coming from city/town $c_i$ at each time step.

• $\mu(t)$: This denotes the hospital service rate at time step $t$.

• **rule:** This represents how the hospital adjust the service rate with respect to the accumulated patient arrivals. The specific rule will be formally described in the next subsection.

• $w(\tau)$: This records the wait time information of hospital $h_j$ in a past time period, which will be released in environment $E$. In this work, the wait time information $w$ released at time round $\tau$ is the average median wait time for the past time round $\tau - 1$. 

146
queue: This records the information about the queue that includes all the patient entities waiting for cardiac surgery services at each time step.

7.3.4 Designing Behavioral Rules

(1) Behavioral Rules for Patients Selecting Hospitals

Based on the literature review and the analysis of variable relationships presented in Chapters 3 and 4, we identify stylized facts regarding the effects of key factors that influence patient-GP mutual decisions for hospital selection and the variations of patient arrivals in hospitals.

- **Stylized fact 1**: The probability that patients select a hospital is exponentially and inversely related to the distance between their homes and a hospital [19].

- **Stylized fact 2**: Patients usually prefer to visit a hospital that is resourceful in terms of personnel (e.g., physicians) and facilities (e.g., ORs) [222, 223, 224]. Hospital resourcefulness and the number of patient arrivals are therefore positively correlated [29].

- **Stylized fact 3**: Patients usually prefer to visit a hospital with shorter wait times [12, 34, 225]. However, a large proportion of patients, especially the elderly, may not have access to wait time information or are less likely to consider the wait times when they select hospitals [34].

Based on the stylized facts, we develop two specific behavioral rules, i.e., a DHW rule and a DH rule, to model how patients choose a hospital. The two behavioral rules are our assumptions in this work, which are defined below.

**Definition 7.6 (DHW rule)**: DHW stands for distance, hospital resourcefulness, and wait times. This rule represents how a patient residing in the location \( c_i \) (\( \forall c_i \in C \)) estimates the arrival probability \( a_{ij} \) for a hospital \( h_j \) (\( \forall h_j \in H \)), using the distance \( d_{ij} \), hospital resourcefulness \( r_j \), and released wait time information \( w_j(\tau) \) at time \( \tau \). The hospital selection probability for a hospital \( h_j \) is calculated by:
\[ a_{ij} = f(d_{ij}) \ast f(r_j) \ast f(w_j(\tau)) \]

\[
\begin{align*}
 f(d_{ij}) &= \frac{d'_{ij}}{\sum_{h_k \in H} d'_{ik}} \\
 d'_{ij} &= \frac{\sum_{h_k \in H} d^d_{ik}}{d^d_{ij}} \\
 f(r_j) &= \frac{r_j^{\alpha_r}}{\sum_{h_k \in H} r_k^{\alpha_r}} \\
 f(w_j(\tau)) &= \frac{w_j^{\alpha_w}(\tau)}{\sum_{h_k \in H} w_j^{\alpha_w}(\tau)}.
\end{align*}
\] (7.4)

Where \( \alpha_d (\alpha_d \in [0, 1]) \), \( \alpha_r (\alpha_r \in [0, 1]) \), and \( \alpha_w (\alpha_w \in [0, 1]) \) are exponential parameters indicating the sensitivity of patients to the factors of distance, hospital resourcefulness, and wait times, respectively.

**Definition 7.7 (DH rule):** DH stands for distance and hospital resourcefulness. This rule represents how a patient chooses a hospital \( h_j \) with respect to the distance \( d_{ij} \) and hospital resourcefulness \( r_j \). The hospital selection probability is calculated by:

\[
\begin{align*}
 a_{ij} &= f(d_{ij}) \ast f(r_j) \\
 f(d_{ij}) &= \frac{d'_{ij}}{\sum_{h_k \in H} d'_{ik}} \\
 d'_{ij} &= \frac{\sum_{h_k \in H} d^d_{ik}}{d^d_{ij}} \\
 f(r_j) &= \frac{r_j^{\alpha_r}}{\sum_{h_k \in H} r_k^{\alpha_r}}.
\end{align*}
\] (7.5)

(2) A Behavioral Rule for Hospitals to Adjust Their Service Rates

Hospitals may periodically change their service rates to adapt to unpredictable patient arrivals. For instance, as shown in Figure 7.11, changes in the throughput,
which represents the actual serviced numbers of patients, follows approximately the same pattern as changes in the patient arrivals in cardiac surgery services in Ontario. The correlation coefficient between the throughput and patient arrivals is 0.896 \((p < 0.0001)\), implying that the service rate of a hospital may vary in accordance with the changes in patient arrivals. We therefore define an S rule for hospitals to adjust their service rates by assuming that service rate of a hospital and the queue length (representing the accumulated patient arrivals at present) is positively and linearly related. The definition of the S rule is given as below.

**Definition 7.8 (S rule):** \( S \) stands for service rate adjustment. This rule represents how a hospital \( h_j \) (\( \forall h_j \in H \)) changes its service rate \( \mu_j(\tilde{T}) \) in response to the aggregated patient arrivals at the past time \( \tilde{T} - 1 \). The service rate adjustment is calculated by:

\[
\mu_j(\tilde{T}) = \bar{\mu}_j \left( a_j \frac{A_j(\tilde{T} - 1)}{\bar{A}_j} + b_j \right),
\]  

(7.6) where \( \tilde{T} \) is the time that hospitals adjust their service rate (usually one week in Ontario [10]); \( \mu_j(\tilde{T}) \) is the service rate of a hospital \( h_j \) at time \( \tilde{T} \); \( \bar{\mu}_j \) is the average service rate of a hospital \( h_j \); \( A_j(\tilde{T} - 1) \) is the aggregated patient arrivals at the time \( \tilde{T} - 1 \); \( \bar{A}_j \) is the average patient arrivals at a hospital \( h_j \); and \( a_j \) and \( b_j \) are two adjustment parameters.

### 7.4 Simulation-Based Experiments

In this section, we conduct simulations based on our AOC-CSS model, aiming to understand the observed tempo-spatial patterns in wait times in the cardiac care services.
Figure 7.11: The number of patient arrivals versus the number of treated cases in the cardiac surgery service in Ontario, Canada, between January 2005 and December 2006.

7.4.1 Experimental Settings

The parameters in the AOC-CSS model are initialized using publicly available data. The CCN published monthly statistical reports on cardiac surgery service utilization in Ontario hospitals between January 2005 and December 2006. The average number of treated cases, the median wait time, and the queue length in a month for each hospital are reported. Therefore, the service rate $\mu_j$ for a hospital $h_j$ can be approximated as the average number of served cases in a day. The arrival rate for each patient type in the city/town $c_i$ can be approximated by:

$$\sum_{k \in K} GP_i \lambda_k = s_i \ast m_i,$$

(7.7)

where $s_i$ is the patient-generation probability, i.e., the probability that a person in the city/town $c_i$ is a patient who needs a cardiac surgery service, and $m_i$ is the size of the total population in the city/town $c_i$. The parameter $s_i$ represents the
heterogeneity of the city/town $c_i$ in producing a patient population requiring cardiac surgery services with respect to its demographics and socioeconomic factors. The patient-generation probabilities for the cities and towns in each LHIN can be inferred from [24]. The total population $m_i$ for each city/town is gathered from the 2006 Canada Census data$^4$.

As seasonal weather is an important contributing factor influencing patient arrivals [22], the arrival rate is adjusted seasonally in our simulation. The patient arrival rate is approximately 15% lower in the warm season (from May to October in Ontario) than in the cold season (from January to April and from November to December in Ontario), according to the reported CCN data.

Near 20% of patients consider wait times when they select a hospital [34]. Therefore, we assume that the probability that a patient considers the factor of wait times when selecting a hospital is relatively small and we set the probability $P_r = 0.2$ in our simulations.

According to the practice, patients are categorized into two types, urgent and non-urgent, i.e., $K = 2$. Following the data reported in [24, p.71], the arrival rate of urgent patients versus that of non-urgent patients is set to 0.23:0.77. Urgent patients have a higher priority for receiving cardiac surgery services than non-urgent patients.

The values of the exponential parameters ($\alpha_d$, $\alpha_r$, and $\alpha_w$) are estimated using the spatial pattern of real patient flows (shown in Figure 7.6). Based on our experiments, we found that the mean and standard deviation of absolute errors have relatively small values when $\alpha_d = 4$, $\alpha_r = 1$, and $\alpha_w = 1$. Here, the absolute error is defined as $|e_{ij}| = |\hat{a}_{ij} - \hat{a}'_{ij}|$, where $e_{ij}$ is the error between the actual percentage of patients residing in LHIN $l_i$ who attend hospitals in LHIN $l_j$ from 2007 to 2008 in Ontario, and that obtained from our simulations.

We run our simulations over two years, so that the simulated data can be directly compared to the observed real-world data. At each time step, the simulation runs 1,000 times and generates an average number of patient for each city/town.

$^4$http://estat.statcan.gc.ca/cgi-win/CNSMCGI.EXE?Lang=E&C91SubDir=ESTAT&DBSelect=FSA06IN
7.4.2 Statistical Regularities in Patient Arrivals and Wait Times

In this section, we examine the statistical regularities in patient arrivals and wait times in our synthetic cardiac care system. Figure 7.12 compares the distribution of the variations in patient arrivals in the real world (represented by the squares in the figure) and the distribution obtained from the simulation (represented by the stars in the figure). The simulation approximately reproduces the shape of the distribution of observed patient-arrival variations, shown in Figure 7.12. The observed patient-arrival variations have a mean of 0.0004 and a standard deviation of 0.226, whereas the simulated patient-arrival variations have a mean of 0.0013 and a standard deviation of 0.232.

![Figure 7.12: Distributions of the variations in simulated and observed patient arrivals in cardiac surgery services. SD: standard deviation.](image)

The relative entropy or the Kullback-Leibler (KL) divergence is a measure of the difference between two probability distributions [227]. The KL divergence of the statistical distribution of simulated patient-arrival variations from that of real-world patient-arrival variations is 0.1398. The small value of the KL divergence implies that the distribution of patient-arrival variations obtained from the simulation may
Figure 7.13: Distribution of simulated absolute wait time variations (by month) in cardiac surgery services. The distribution follows a power law with a power of -1.47 (power-law test based on Clauset’s method [215]: $p < 0.1$; linear fitness (red line): $p < 0.0001$; standard deviation $SD = 0.183$).

Figure 7.13 presents the statistical distribution of absolute variations in the median wait time obtained from our simulation. The absolute variations in the median wait time follow a power-law distribution with a power of -1.47 (linear fitness: $p < 0.0001$). The fitness of the power-law distribution is tested using the Clauset method [212] (power-law test: $p < 0.1$). This distribution indicates that the synthetic cardiac surgery service is self-organizing in terms of its wait times.

Figure 7.14 compares the statistical distribution of absolute variations in the median wait time obtained from our simulation to the distribution of the observed data. The KL divergence of the distribution of the simulated absolute wait-time variations (represented by stars in the figure) from that of the observed absolute wait-time variations (represented by squares in the figure) is 0.1227. The small value of the KL divergence implies that the two distributions are similar.
Figure 7.14: Distributions of simulated and observed wait time variations in the cardiac surgery service.

### 7.4.3 Patient-Attraction and Patient-Distribution Degrees of LHINs

Figure 7.15 compares the observed and simulated distributions of LHINs’ patient-attraction degrees and patient-distribution degrees. The simulated patient-attraction and patient-distribution degrees for each LHIN are approximately the same as the observed degrees, except for LHIN 12, which has a lower simulated patient-distribution degree than that observed in the real world. In the simulation, most patients living in LHIN 12 select the hospital in LHIN 8 because it has the shortest driving time (0.6 hour). However, in the real world, although LHIN 6 and 7 are not next to LHIN 12 and have longer driving times (1.2 hour and 1.1 hour, respectively), approximately 25% of patients who live in LHIN 12 visit the four hospitals (Trillium Health Centre, St. Michael’s Hospital, Sunnybrook Health Sciences Centre, and University Health Network) in LHIN 6 and 7, as these hospitals have good resources and the driving time between homes and these hospitals are short enough to be acceptable.

The simulated distributions in Figure 7.15 exhibit low- and
Figure 7.15: Comparison of the simulated and observed distributions of LHINs’ patient-attraction and patient-distribution degrees.

High-patient-attraction groups, and low- and high-patient-distribution groups that are almost the same as the groups exhibited by the observed distributions, shown in Figure 7.7. LHINs 5-9 have obviously larger patient-distribution degrees and thus form a high-patient-distribution group, whereas LHINs 1-4 and 10-13 have less patients travelling to other LHINs for cardiac surgery services and thus fall into the low-patient-distribution group. Similarly to the observed distributions, LHINs 6-8 exhibit higher attraction degrees in the simulation and thus form a high-patient-attraction group, whereas the other LHINs (LHINs 1-5 and 9-14) fall into the low-patient-attraction group.

7.4.4 Tempo-Spatial Patterns in Patient Arrivals and Wait Times

(1) The Dynamics of Patient Arrivals in Each Hospital

Figures 7.16 and 7.17 compare the observed and simulated temporal patterns in patient arrivals for each hospital and show that our AOC-CSS model is able to approximately reproduce the observed temporal patterns in patient arrivals, as the correlation coefficient $R$ of the simulated and observed patient arrival variations for
Each hospital is positive.

(2) The Dynamics of Wait Times in Each Hospital

Figures 7.18 and 7.19 compare the observed and simulated temporal patterns of the median wait times for each hospital and show that our AOC-CSS model is able to approximately reproduce the observed temporal pattern of median wait times, as the correlation coefficient $R$ of the simulated and observed patient arrival variations for each hospital is positive.

7.5 Discussion

7.5.1 Explaining the Underlying Causes of Tempo-Spatial Patterns

Based on our AOC-CSS model and simulation-based experiments, we are able to characterize the tempo-spatial patterns in patient arrivals and wait times as observed in real-world cardiac surgery services. These patterns are partially due to the local feedback loop between patient arrivals and hospital wait times, shown in Figure 7.10.

Let us take the city of Brampton, Ontario, as an example to illustrate the self-organizing process at an individual level. The four hospitals nearest to Brampton that offer cardiac surgery services are Trillium Health Centre (H4), St. Michael’s Hospital (H5), Sunnybrook Hospital (H6), and University Health Network (H7). The average driving times for patients living in Brampton to travel to these hospitals are less than 0.7 hour. Figure 7.20 presents the dynamically changing preferences of patients residing in Brampton for the four hospitals and shows that patients living in Brampton generally prefer H7, because the driving distances from Brampton to the four hospitals are almost the same, varying between 0.5 hour and 0.7 hour, and H7 has more physicians than the other three hospitals. As the values for the factors of driving distance and hospital resourcefulness are not changed during the simulation,
Figure 7.16: The observed and simulated temporal patterns in patient arrivals in hospitals H1-H6.
Figure 7.17: The observed and simulated temporal patterns in patient arrivals in hospitals H7-H11.
Figure 7.18: The observed and simulated temporal patterns in wait times in hospitals H1-H6
Figure 7.19: The observed and simulated temporal patterns in wait times in hospitals H7-H11.
the changing wait times for the four hospitals are the only cause of the dynamically changing arrival probabilities.

For instance, Figure 7.20(d) shows that in the first two months, the arrival probabilities for patients living in Brampton for H7 are high, because the wait times in this hospital are short, at approximately 22 days. Due to the high arrival probabilities in the first two months, more patients may prefer to visit H7 than the other three hospitals, which will in turn result in longer wait times in H7. The wait time information for H7 is then released into the environment and is used by patients when they make hospital selection decisions in the third month. As a result, the arrival probability of patients living in Brampton for H7 in the third month will decrease. This self-regulating process is initiated by autonomous patient/GP entities according to their hospital selection behavioral rules and incorporates the feedback loop between wait times and hospital selection behavior, potentially accounting for the observed self-organized tempo-spatial patterns at a systems level.
Figure 7.20: The dynamically changing preferences of patients residing in the city of Brampton (in LHIN 5) to the four neighboring hospitals. (a): H4, Trillium Health Centre; (b): H5, St. Michael’s Hospital; (c): H6, Sunnybrook Hospital; (d): H7, University Health Network. The shaded areas in this figure represent the warm seasons in Ontario, Canada.
Figure 7.20 also shows that the trends of the changes in arrival probabilities for the four hospitals are complementary. The increase in arrival probabilities to some of the hospitals in some months therefore accompanies the decrease in arrival probabilities to other hospitals. Due to the differences in the wait times in the four hospitals, a few patients may therefore transfer between the four hospitals to avoid a long wait. For instance, in the first warm season (from month 3 to month 8), the arrival probabilities for H4 and H6 increase because their reference wait times are less than 20 days, whereas the arrival probabilities for H5 and H7 decrease because their wait times are much longer than 20 days. It should be noted that although the arrival probabilities for H4 and H6 increase, the wait times in all four hospitals decrease in the first warm season. The number of patient arrivals in the warm season is smaller than in the cold season. As more patients may be willing to travel to H4 and H6 in the first warm season, the accumulated patient arrivals in the first warm season may result in the increase in wait times in the initial several months in the second cold season (from month 9 to month 12), which will in turn reduce the arrival probabilities for the two hospitals. With the same analysis process described above, we can explain the variations in the arrival probabilities and wait times for the four hospitals in the subsequent months.

7.5.2 Sensitivity Analysis

To investigate the sensitivity of our results, we now discuss the statistical distributions of the median wait times with respect to different time scales for calculating the variations in wait times; and different probabilities (i.e., $P_r$) that a patient takes the wait time information into account when making hospital selection decisions.

(1) Wait Time Variations at Different Time Scales

Figures 7.21 and 7.22 show the statistical distributions of absolute wait time variations calculated by week and by half-month, respectively. We use the method
developed by Clauset et al. [212] to test whether our simulated data follows the power law. We find that the absolute wait time variations presented in the two figures both fit a power-law distribution (power-law test: \( p < 0.1 \)). The power of the statistical distribution calculated by week is -2.19 and calculated by half-month is -1.86, suggesting that absolute wait time variations in different time scales are able to represent the self-organizing property of the cardiac care system in terms of wait times, such as by week, as shown in Figure 7.21, by half-month, as shown in Figure 7.22, and by month, as shown in Figure 7.13.

![Figure 7.21: Distribution of simulated absolute wait time variations (calculated by week) in cardiac surgery services. The distribution follows a power law with a power of -2.19 (power-law test based on Clauset’s method [215]: \( p < 0.1 \); linear fitness (red line): \( p < 0.0001 \); standard deviation \( SD = 0.331 \)).](image)

(2) The Probability for Selecting DHW Rule, \( P_r \)

Figure 7.23 shows the distributions of absolute wait time variations (calculated by month) in cardiac surgery services with respect to different probabilities that a patient considers wait times when choosing a hospital, \( P_r \). Table 7.1 presents the corresponding \( p \)-values of power-law tests with respect to various \( P_r \) based on
Figure 7.22: Distribution of simulated absolute wait time variations (calculated by half-month) in cardiac surgery services. The distribution follows a power law with a power of -1.86 (power-law test based on Clauset’s method [215]: $p < 0.1$; linear fitness (red line): $p < 0.001$; standard deviation $SD = 0.38$).

Table 7.1: The $p$-values of power-law tests for distributions of absolute wait time variations with respect to different $P_r$

<table>
<thead>
<tr>
<th>$P_r$</th>
<th>1</th>
<th>0.75</th>
<th>0.5</th>
<th>0.25</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.10</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Note: If $p \leq 0.1$ as suggested by Clauset et al. [215], the data for the power-law fitness test follows a power-law distribution.

Clauset’s method [212]. According to Figure 7.23 and Table 7.1, when there are no wait time-sensitive patients (i.e., $P_r=0$) who take into account the wait time information when choosing hospitals, the distribution of absolute wait time variations does not follow a power-law distribution, as the power-law test is not significant ($p = 0.13$). If all of the patients select hospitals without considering the wait time information, the feedback loop between the patient hospital selection behavior and wait times is absent. Thus, patient arrivals cannot adapt to the dynamically changing wait times in hospitals.
Figure 7.23: Distributions of absolute wait time variations (by month) in cardiac surgery services with respect to different $P_r$.

According to Figure 7.23 and Table 7.1, when there is a relatively small probability that a patient considers wait times when choosing a hospital, e.g., $P_r = 0.25$, the distribution of absolute variations in the median wait time follows a power-law distribution ($p = 0.10$), suggesting that the system is self-regulating due to the wait time-sensitive patients who select different hospitals in accordance with the variations in wait times in each hospital.

However, when $P_r$ becomes large, for instance, 0.5, 0.75, or 1, as shown in Figure 7.23 and Table 7.1, the distributions of absolute wait time variations do not follow power-law distributions. The $p$-values of the power-law tests are all larger than 0.1. A large number of wait time-sensitive patients may therefore not result in a self-regulating healthcare service system, as the patient arrivals for each hospital may fluctuate highly if more patients are sensitive to the wait time information when they select hospitals. This can be observed in Figure 7.23, which shows that larger $P_r$ values result in larger variations in absolute wait time variations.
7.6 Summary

In this chapter, we used a behavior-based autonomy-oriented modeling method to characterize the tempo-spatial patterns in a cardiac care system from a complex self-organizing systems perspective. We described three types of entities, patients, GPs, and hospitals, and the environment that they reside in and access information from. Based on the identified major impact factors of distance, hospital resourcefulness, and wait times, and their interaction relationships and local feedback loops, we derived specific behavioral rules for wait time-sensitive and wait time-insensitive patients to make mutual decisions with their GPs on hospital selection. We also designed a specific behavioral rule for hospitals to adjust their service rates with respect to the waiting patients. Through simulation-based experiments, we observed that the constructed white-box AOC-CSS model produces tempo-spatial patterns that are approximately similar to those observed in the real-world cardiac surgery system. The patient-GP mutual hospital selection behavior and its relationship with hospital wait times may therefore account for self-regulating service utilization. The study also revealed that the behavior-based autonomy-oriented modeling method provides a potentially effective means for explaining the self-organized regularities and investigating emergent phenomena in complex healthcare systems.
Chapter 8

Conclusions and Future Work

This chapter concludes the thesis by summarizing the main contributions and suggesting directions for future work.

8.1 Summary

In this thesis, we addressed the challenging problem of understanding the performance of healthcare services by investigating the causes of the dynamics of and the relevant tempo-spatial patterns in wait times from a self-organizing systems perspective. We developed a data-driven complex systems modeling approach that: (1) uses SEM-based analysis to explore the direct, indirect, and moderating effects of demand and supply factors on wait times; uses (2) integrated projection to estimate the changes in service utilization and performance with respect to demographic shifts; (3) designs and evaluates service management strategies to improve the utilization of healthcare service resources and wait time management; and (4) uses behavior-based autonomy-oriented modeling to characterize the emergent tempo-spatial patterns in wait times from a self-organizing systems perspective.

The uniqueness of this approach lies in the following aspects.

1. The capacity for exploring the direct, indirect, and moderating effects of demand and supply factors on wait times using an SEM-based analysis.
Unlike existing studies that use traditional multivariate data analysis methods, such as regression, to investigate pairwise relationships between observed variables, the \((SEM)\)-based analysis method enables us to construct LVs and to explore the direct, indirect, and moderating relationships between multiple observed variables and LVs.

2. The capacity for estimating and demonstrating the changes of service utilization and performance with respect to demographic shifts using the integrated projection method.

Unlike existing studies that estimate the utilization and performance of specific healthcare services using pairwise relationships between certain variables, or queuing model-based discrete event simulations, the proposed integrated projection method is able to estimate service utilization and service performance by taking into account the complex effects exerted by multiple observed and latent predictors, and to represent the dynamically changing process of the estimated service utilization and service performance over time.

3. The capacity for designing and evaluating adaptive service management strategies for improving the utilization of healthcare service resources and wait time management.

Unlike existing studies, the adaptive strategy for time block allocation in ORs proposed and evaluated here is designed from a complex systems perspective. The proposed adaptive strategy can adaptively allocate OR time blocks in response to unpredictable changes in patient arrivals by incorporating historical information. The strategy therefore addresses the issue of long waits due to ineffective use of service resources.

4. The capacity for characterizing the emergent tempo-spatial patterns in wait times using behavior-based autonomy-oriented modeling.

Unlike existing methods for modeling and simulating the dynamic behavior
of a system, the proposed method models healthcare services from a self-organizing systems perspective, with a special focus on individuals’ autonomous behavior, their interactions, and local feedback loops with respect to the effects of multiple factors. Based on the modeled synthetic healthcare services, working mechanisms that account for the self-organization process and the observed emergent tempo-spatial patterns in wait times may be discovered.

The proposed methods were implemented to investigate the dynamics of and relevant patterns in wait times in the cardiac surgery services in Ontario, Canada. The major findings and resulting suggestions for healthcare administrators to improve healthcare service management are summarized below.

1. As presented in Chapter 3, we have found that service accessibility and the education profile alleviate the effects of the population size and/or age profile on service utilization, by using the SEM-based analysis method. This finding suggests that efforts to reduce regional disparities in healthcare service utilization should take into account the interactions between geodemographic factors, such as service accessibility and education. When resources are allocated to a particular healthcare service in one area, the geographic distribution of the same services in neighboring areas should be considered, as patients may be willing to use services in areas not far from their residence.

2. As described in Chapter 4, we have found that the characteristics of a CU have direct and indirect effects on the wait times in an SU, by using the SEM-based analysis method. The demand for and wait times in the preceding unit are good predictors for the wait times in subsequent units. Cross-unit effects therefore must be considered when attempting to alleviate the wait times in a healthcare service system. Further, different patient risk profiles may affect wait times in different ways (e.g., positive or negative effects) within the SU.
Wait time management should therefore carefully consider the relationship between priority triage and risk stratification, especially for cardiac surgery.

3. As demonstrated in Chapter 5, we have projected the regional service utilization and wait times for cardiac surgery services in Ontario in 2010 and 2011, based on statistics for these services between 2005 and 2007 and the changing trends in demographics between 2006 and 2011. The evaluated projection results show that the proposed integrated method may be useful for envisioning the changes in and dynamics of service utilization and service performance with respect to demographic shifts.

4. As shown in Chapter 6, we have proposed and evaluated an adaptive strategy for allocating OR time blocks. We have also applied the adaptive strategy to a specific queuing model that was built on the general perioperative process of the CS-ORs in the HHSC. We have found that our adaptive strategy is able to efficiently regulate the OR time block reservations in accordance with dynamic changes in patient arrivals. Our experimental findings suggest that frequently adjusting the allocation of OR time blocks (i.e., once a month) will help to improve OR utilization.

5. As described in Chapter 7, we have modeled and simulated cardiac surgery services using a behavior-based autonomy-oriented modeling method. Through simulation-based experiments, we observed that the constructed white-box AOC-CSS model produces global-level regularities similar to those found in the real-world cardiac surgery system. The patient-GP mutual hospital selection behavior and its relationship with hospital wait times may therefore account for the self-regulation of wait times in cardiac surgery services.

The data-driven complex systems modeling approach presented in this thesis can be generalized to investigate the causes and corresponding dynamics of performance indicators besides wait times in different healthcare systems. For instance, we may
investigate how treatment costs and wait times affect patients’ decisions on selecting public hospitals or private hospitals in Hong Kong. To study this problem, we may first use the method of *SEM-based analysis* to develop a specific SEM model that includes the factors of geographic profiles in Hong Kong and the characteristics of the public/private hospitals. Based on the aggregated data about Hong Kong, we can test the factors and the different effects on the patient arrivals to public hospitals and private hospitals, respectively. Then, using the method of *behavior-based autonomy-oriented modeling*, we can model the behaviors and interactions of heterogeneous patients and hospitals to represent the real situations in Hong Kong. Furthermore, with the identified relationships between the determinants and patient arrivals, we can use the *integrated projection* method to estimate the changes of service utilization in a specific hospital in Hong Kong if some of the determinants are changed.

In summary, the data-driven complex systems modeling approach presented in this thesis provides a practically useful means for understanding the dynamics of and relevant emergent patterns in specific healthcare services, for estimating the changes in service performance, and for designing better strategies for improving service management. The proposed methods, demonstrated analytical processes, and corresponding findings offer policy makers, researchers, and healthcare administrators insights for improving wait time management in a specific healthcare service system.

### 8.2 Future Work

There are several directions for extending this research.

1. **Heterogeneities of physicians**

   The work presented in this thesis assumes that all physicians in hospitals are homogeneous in serving patients. Based on this assumption, specific queuing models have been developed to characterize the behavior of hospitals. As
physicians may differ in how they prioritize and serve patients [204], the allocated time blocks for serving patients [228], and their medical skills [228], future research could take into account the heterogeneity in physicians. The resulting extended hospital models may better represent the actual operation of healthcare services. Furthermore, modeling heterogeneous physicians may help to design more practical strategies for improving service management behavior.

2. Social influence on patients’ hospital selection behavior

In Chapter 7, the proposed AOC-CSS model does not take into account how opinions and experiences of socially connected people affect patients’ hospital selection decisions. Prior patients’ experiences with specific service providers and physicians may spread through a social network, and thus affect the hospital selection decisions of subsequent patients. The influence of socially connected people, which is referred to as social influence in the literature, plays a significant role in patients’ choice of hospitals [229]. It therefore would be valuable to extend the current AOC-CSS model by incorporating the effects of social influence. By doing so, we may thus develop a more realistic model for characterizing patients’ hospital selection behavior. The extended model may also enable us to evaluate the effects of patients’ experience and social interactions on the emergent tempo-spatial patterns in service utilization and wait times.

3. Perceived hospital reputation on patients’ hospital selection behavior

In Chapter 7, we have investigated how patients select hospitals with respect to the factors of distance, resourcefulness, and wait times, which are identified as key impact factors based on the SEM-based analysis and the literature review. However, the perceived hospital reputation, an unobserved factor that covers patients’ perceptions of multiple dimensions, such as hospital resourcefulness, physicians’ medical skill, and service outcome, is another important factor influencing patients’ choices of a hospital [230, 231]. In the future work, it
would be interesting for us to consider the perceived hospital reputation into our AOC-CSS model, so as to better represent the real-world situations and evaluate the effects of hospital reputation on the dynamics of patient arrivals to different hospitals.

4. Service utilization projection and resource allocation optimization

The method presented in Chapter 5 for projecting service utilization and service performance in the future is based on the assumption that the relationships between the considered impact factors, service utilization, and service performance are homogeneous across all hospitals and will not change in the estimation period. This assumption can be relaxed by carrying out projections based on specific individual-based models developed with the behavior-based autonomy-oriented modeling method. Most of the current healthcare resource allocation strategies do not take into account the response behavior of patients with respect to new allocated resources in specific hospitals [130], which potentially result in substantial differences between the estimated need and the real need for healthcare resources [146]. Our behavior-based autonomy-oriented modeling method can develop models to characterize such behavior and thus support policy makers and healthcare administrators in designing better resource allocation strategies and evaluating their effectiveness in improving service performance.
Bibliography


[7] Department of Health and Human Services Data Council, Centers for Disease Control and Prevention, National Center for Health Statistics, National Committee on Vital and Health Statistics.


196


where to go for major surgery? Telephone interview survey. BMJ, 331:821, 
2005.

An experimental study. In Proceedings of the Second Advances in Health 

Cardiac Surgical Service. Thesis of Masters of Applied Science, University of 
Toronto, 2006.

strategies for improving operating room efficiency at academic institutions. 


[213] S. Kauffman. At Home in the Universe: The Search for the Laws of Self-

(1).pdf.


Curriculum Vitae

Academic qualifications of the thesis author, Ms. TAO Li:

- Received the degree of Bachelor of Engineering from Southwest Normal University, June 2003.
- Received the degree of Master of Engineering from Southwest University, June 2007.

Publication lists:


February 2014