Analytical Models for Specialist Care in Rural Areas

Michael G. Klein

Doctor of Philosophy

Desautels Faculty of Management

McGill University

Montreal,Quebec

December 2016

A thesis submitted to McGill University
in partial fulfillment of the requirements for the degree of
Ph.D. in Healthcare Operations Mangement

DEDICATION

To Azusa and Tito, for their ongoing support and encouragement.

ACKNOWLEDGEMENTS

Thank you to my advisor Professor Vedat Verter for his guidance in my pursuit to make this contribution to the important field of healthcare operations research. While our initial goal was to focus on emergency department operations management problems applied in the urban setting of Montreal, he continued to serve as my Ph.D. advisor and supported my research ideas to study rural healthcare problems instead. He also agreed to continue to serve as my Ph.D. advisor remotely and encouraged me to follow my desire to investigate the understudied topic of Specialist Care. I am also thankful for the financial support received from Professor Verter's Natural Sciences and Engineering Research Council (NSERC) Collaborative Research and Training Experience (CREATE) Program Grant on Healthcare Operations and Information Management.

After leaving Montreal, I moved to Nova Scotia and looked to study rural healthcare problems there. I searched for appropriate research collaborators and among them was Dr. Brian Moses, Chief of Internal Medicine in Yarmouth Regional Hospital (YRH). I thank Dr. Moses for his interest, time, and ongoing support of both of my Ph.D. research projects and for serving on my Ph.D. committee. I also thank Professor Jean-François Cordeau for serving on my Ph.D. committee. He provided insightful comments, questions and suggestions.

Thank you as well to Dr. Hughie Fraser, Department Head of Internal Medicine in South Shore Regional Hospital (SSRH) for his collaboration and support. I also thank all of the other Internal Medicine specialists from YRH and SSRH for allowing me to observe their work and for helping me gather additional data. Thank you to the decision support groups from YRH and SSRH who provided hospital information systems data and the Nova Scotia Renal Program Management for answering my questions and for providing dialysis cost information.

Finally, I thank my wife and son who had a challenging journey during the years of my Ph.D. studies. Without their ongoing love and support, this dissertation would not have been possible.

CONTRIBUTION OF AUTHORS

Three manuscripts are based on this thesis:

1. Verter V, **Klein MG**. Operations Research for Emergency Care: A Review. Invited for publication in *Production and Operations Management*

   The first author received an invitation to provide a healthcare review paper for publication in *Production and Operations Management*. After the first author suggested to conduct a systematic review of the Emergency Department Operations Management literature, the second author (student) completed the systematic review of the literature and discussed the most relevant studies with the first author. The second author prepared Chapter 2, providing a detailed review of the literature. Chapter 2 represents the first draft of the review paper to be refined by both authors before submission for publication.

2. **Klein MG**, Verter V, Fraser HF, Moses BG. Specialist Care in Rural Hospitals: From Emergency Department Consultation to Inpatient Ward Discharge. Target: *Manufacturing and Service Operations Management*

   The first author (student) was the main person responsible for conducting the research. After identifying and defining the research problem and developing the modeling framework, the first author established collaborations with the third and fourth authors for this study. The third and fourth authors served as supervising investigators for the ethical conduct of the first author in the South Shore Regional Hospital and Yarmouth Regional Hospital respectively. The third and fourth authors facilitated the recruitment of all Internal Medicine specialists (Internists) for Internist data collection purposes by informing the other specialists of the study in department meetings and providing the Internal Medicine on call schedules. The first author collected data by observing all

Internists on call in both hospitals, and by asking the Internists to complete data collection forms developed by the first author. The first author also worked with hospital decision support specialists to obtain data available in hospital information systems. Throughout the study, the first author worked with guidance from the second author. The second author reviewed the research problem and modeling framework and helped design the experiments for computational analysis. The first author developed computer programs and simulation models to complete the computational analysis. After the first author prepared a draft of the manuscript, the second author provided suggestions to help improve the manuscript. The third and fourth authors also shared medical and hospital management expertise and reviewed the manuscript.

3. **Klein MG**, Verter V, Moses BG. Dialysis Facility Network Design. Target: *Journal of Operations Management*

   The first author (student) was the main person responsible for conducting the research. The research problem was identified by the first and third authors. The first author then developed the mathematical model and reviewed the research problem and model with the second author. The third author served as supervising investigator for the ethical conduct of the first author in the Yarmouth Regional Hospital. The first author developed the patient survey and the second and third authors reviewed the survey. The first and third authors conducted the patient survey and the first author de-identified patient information and recorded patient responses. The first author developed the methodology to generate province-wide patient residence data and calculated travel times to potential facility locations. After the third author contacted renal program management in Halifax, the first author obtained cost data from renal program management. Given the second author's expertise from a methodology perspective, the second author reviewed the first author's refinements to the mathematical model and provided guidance to the first author. The first author designed the experiments for computational analysis, developed computer programs and applied the

model with the data sources to generate solutions and insights. After the first author reviewed results of the computational experiments with the second author, the second author suggested additional experiments. The first author then prepared Chapter 4 which represents the first draft of the manuscript to be edited and refined by all authors before submission for publication.

ABSTRACT

Healthcare systems rely on specialists to solve many medical problems. In my research, I explore the Operations Management challenges of specialists for acute and chronic patient care. I develop general models and consider the additional challenges of providing Specialist Care in Rural Areas. This thesis contains two research projects: the first project focuses on acute care while the second project focuses on chronic care.

In the first project, I study the workflow decisions of specialists on call in rural hospitals. Specialists receive consultation requests for new patients in the Emergency Department (ED) and take care of inpatients in hospital wards. Should specialists give priority to ED consultation requests or give priority to inpatient discharges? I propose a stochastic dynamic programming framework for Specialist Care that includes a Single Role Model and a Dual Role Model. In rural hospitals, Internal Medicine Specialists (Internists) typically take on a dual role as Intensive Care Unit (ICU) physician and Internist on call. I apply the proposed modeling framework to data sets developed from two corresponding case studies. One hospital uses the traditional rural approach and the other hospital staffs a separate Internist for the ICU. After observing all Internists working on call, I work with the two physician groups to obtain ED consultation and inpatient care data and combine it with hospital information system data. I find that an early inpatient discharge policy suggested by current guidelines (i.e. discharge inpatients by 11:00AM) is not always a good strategy. In hospitals with the common challenge of ED congestion and boarding, specialists should sometimes prioritize inpatient discharges and other times give priority to ED consultations. I find optimal policies that have different forms throughout the day with boarding thresholds and end-of-horizon effects.

For the second project, I study the chronic care problem of Dialysis Facility Network Design. Kidney specialists treat chronic kidney failure with dialysis until transplant or death.

Patients travel to in-centre or satellite hemodialysis (HD) facilities for each four hour treatment, three times per week or participate in home peritoneal dialysis (PD) or home HD. The travel burden for patients in rural areas can be greater than one hour in each direction. Regardless of the travel burden, some patients will always opt to go to an in-centre or satellite facility, while others will always opt for home dialysis. For many, the choice will vary depending on the location of available facilities. I develop a mathematical model for the Dialysis Facility Network Design Problem, considering the impact of travel distance on patient choice for dialysis mode. Using dialysis patient surveys, I obtain patient preference data for facility-based or home dialysis in Nova Scotia. The model also incorporates the challenges of capacity management and budget constraints required to find a feasible solution. I apply the model to identify the best network of facilities to reduce travel time for those most in need, improving the welfare of the dialysis patient population.

RÉSUMÉ

Les systèmes de santé comptent sur les spécialistes pour résoudre de nombreux problèmes médicaux. Dans mes recherches, j'explore les défis de gestion des opérations de spécialistes pour les soins aux patients aigus et chroniques. Je développe des modèles généraux et considère les autres défis de la prestation de soins de spécialistes dans les zones rurales. Cette thèse contient deux projets de recherche: le premier projet se concentre sur les soins actifs alors que le second projet se concentre sur les soins chroniques.

Dans le premier projet, j'étudie les décisions de flux de travail de spécialistes sur appel dans les hôpitaux ruraux. Les spécialistes reçoivent des demandes de consultation pour les nouveaux patients dans le service d'urgence et de prendre soin des patients hospitalisés dans les services hospitaliers. Est-il préférable de donner la priorité à l'urgence ou aux rejets d'hospitalisation? Je propose un cadre de programmation dynamique stochastique pour les soins de spécialistes qui comprend un modèle de rôle unique et un modèle de rôle double. Dans les hôpitaux ruraux, spécialistes en médecine interne (internistes) prennent généralement sur un double rôle de l'unité de soins intensifs médecin et interniste sur appel. Je demande le cadre de modélisation proposé aux ensembles de données mis au point à partir de deux études de cas correspondantes. Un hôpital utilise l'approche traditionnelle rurale et l'autre hôpital utilise un interniste distinct pour l'unité de soins intensifs. Après avoir observé toutes les internistes travaillant sur appel, je travaille avec les deux groupes de médecins pour obtenir des données de consultation urgence et de soins aux patients hospitalisés et les combiner avec les données du système d'information hospitalier. Je trouve que la politique des patients hospitalisés au début de décharge proposé par les lignes directrices actuelles (à savoir les patients hospitalisés à décharge par 11h00) ne sont pas toujours une bonne stratégie. Dans les hôpitaux avec le défi commun de la congestion urgence et l'embarquement, les spécialistes doivent parfois prioriser les rejets d'hospitalisation et d'autres fois donner la priorité aux consultations à l'urgence. Je trouve des politiques optimales qui ont différentes

formes tout au long de la journée avec des seuils d'embarquement et les effets de fin d'horizon. Pour le deuxième projet, j'étudie le problème chronique de soins de l'établissement pour dialyse, *Dialysis Facility Network Design Problem (DFNDP)*. Spécialistes du rein traiter l'insuffisance rénale chronique par dialyse jusqu'à ce que la greffe ou la mort. Les patients se déplacent dans le centre ou l'hémodialyse par satellite d'installations pour chaque traitement de quatre heures, trois fois par semaine ou participer à la dialyse à domicile péritonéale ou l'hémodialyse à la maison. Le fardeau de voyage pour les patients dans les zones rurales peut être plus d'une heure dans chaque direction. Quelle que soit la charge de voyage, certains patients seront toujours choisir d'aller à un centre ou satellite, tandis que d'autres seront toujours opter pour la dialyse à domicile. Pour beaucoup, le choix varie en fonction de l'emplacement des installations disponibles. Je développe un modèle mathématique pour le DFNDP, compte tenu de l'impact de la distance de voyage sur le choix du patient pour le mode de dialyse. Utilisation de dialyse sondages auprès des patients, j'obtenir des données de la préférence du patient pour en établissement ou la dialyse domicile en Nouvelle-Écosse. Le modèle intègre également les défis de contraintes de gestion et budget capacités nécessaires pour trouver une solution réalisable. Je demande le modèle pour identifier le meilleur réseau d'installations pour réduire le temps de voyage pour les personnes les plus dans le besoin, l'amélioration du bien-être de la population de patients de dialyse.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
Introduction to Rural Healthcare

Health disparities are associated with socioeconomic status, measured by variables including education, occupation, income, wealth and place of residence. Such health disparities are widespread, and variations by geographic region suggest that disparities are avoidable (Adler and Rehkopf 2008). In the United States (U.S.), rural residents exercise less, have less nutritional diets, smoke more and are more likely to be in fair or poor health. Compared to urban and suburban populations, rural areas have relatively more elderly people, higher unemployment rates, less education, more poverty and more uninsured. In addition to the usual challenges of chronic disease management and reduced mobility, the rural elderly also face geographic isolation. Even though rural communities have greater need for health care services, they receive much less than their share of health care resources (Ricketts 2000, Rosenthal and Fox 2000, Hartley 2004, Hart et al. 2005).

With life expectancy among the highest of the Organisation for Economic Co-operation and Development (OECD) countries, most Canadians live long healthy lives. However, health disparities by geographic location are also prevalent in Canada. Rural Canadians have less healthy behaviours including higher proportions of smokers, lower consumption of fruits and vegetables and a higher proportion of overweight people than those living in urban areas. Rural residents are also less educated and are more likely to be in poorer socio-economic conditions (DesMeules 2006).

Overall, both rural Americans and rural Canadians have higher mortality rates than urban residents. There are considerably higher death rates due to injury and poisoning, possibly due to certain rural-based industries, such as farming, fishing, logging and mining that have high levels of occupational hazards. The suicide rate among men in the most rural U.S. counties is also much higher than the rate among men in suburban counties. Death

rates from chronic obstructive pulmonary disease (COPD) are higher among men who live in nonmetropolitan U.S. counties, and in southern US, heart disease death rates are highest in rural areas (DesMeules 2006, Eberhardt and Pamuk 2004).

1.1 Differentiating Characteristics of the Rural Context for Healthcare

Distance to a primary care provider (PCP) is an important factor in determining the frequency of patient visits, with greater distance resulting in fewer regular check-ups. Rural patients also have fewer chronic care appointments and less utilization of preventive care (Ricketts 2000, Arcury et al. 2005). For example, in absence of early cancer detection programs, cancer has been diagnosed at more advanced and later stages of disease in rural populations compared to urban populations (Monroe et al. 1991).

It is well known that physicians, especially specialists, are concentrated in urban areas. One quarter of the U.S. population lives in rural areas, but only one eighth of physicians work in rural areas (Ricketts 1999). More than 75% of rural U.S. hospital CEOs have reported physician shortages including family medicine (58.3%), general internal medicine (53.1%), psychiatry (46.6%), general surgery (39.9%), neurology (36.4%), cardiology (35.0%) and obstetrics-gynecology (34.4%). In addition to physician shortages, the three most commonly needed allied health professions were registered nurses (73.5%), physical therapists (61.2%) and pharmacists (51%) (Ricketts, 2000). A study of the supply of physicians confirmed that there is a much lower supply of physicians in rural areas. Rural areas had 5.3 PCPs and 5.4 specialists per 10,000 population compared to 7.8 PCPs and 13.4 specialists in urban areas. Furthermore, rural patients have longer waits to get specialist appointments (Reschovsky and Staiti 2005).

With a short supply of specialists even in urban areas, the staffing problem becomes even more challenging for rural areas. Furthermore, the prevalence of conditions requiring specialty care is increasing, disorders that previously were untreatable have definitive therapy, hip and knee replacements are routine, and patients with diseases such as leukemia and colon cancer survive longer, requiring more care from specialists (Cooper 2002). The gap

is sometimes filled by international medical graduates (IMGs), defined as physicians who graduated from medical school outside of the U.S. and Canada. Overall, IMGs account for approximately one quarter of U.S. physicians, with the remaining 75% graduates from U.S. or Canadian medical schools. IMGs are significantly more likely to practice internal medicine (48.2% vs 34.0%), however, the use of IMGs varies by region. In large rural areas, there are relatively more IMGs compared to the U.S. national percent within states such as Wyoming (30.5%), New Mexico (13.0%) and Iowa (9%). In small rural areas, there are relatively more IMGs compared to the U.S. national percent within states such as Maine (24.6%), Delaware (12.0%) and Kentucky (10.3%). In isolated rural areas, there are relatively more IMGs compared to the U.S. national percent within states such as Montana (42.0%), North Dakota (17.2%) and South Dakota (14.5%) (Thompson et al. 2009).

Recruitment and retention efforts to get physicians to practice in rural communities range from selected rotational experiences to full-time residency training affiliated with urban academic centers. For example, Dalhousie University based in Halifax, Nova Scotia has a Family Medicine program that brings medical students to Yarmouth Regional Hospital, roughly 300 km from Halifax. Such endeavours might help develop familiarity, community, sense of place, and knowledge of a supportive and nurturing rural environment. However, the strongest known influence on rural physician recruitment is a rural upbringing and medical schools have fewer rural-raised students and more urban-raised applicants. Those without a rural upbringing are both less familiar with rural life and also less likely to become engaged in communities. As a result, the majority of medical graduates are less likely to choose rural practice (Hancock et al. 2009, Rosenthal and Fox 2000). For applicants with interest in rural practice, the recruitment and retention efforts of greatest importance are 1) healthcare is a major part of the local economy, 2) the community is a good place for family, 3) doctors are well-respected and supported and 4) people in the community are friendly and supportive of each other (Ricketts 2000).

In addition to IMGs, rural communities without enough PCPs sometimes bring in nurse practitioners (NPs) and physician assistants to fill the gap for outpatient care and hospitalists for inpatient care (Cooper 2002). In Nova Scotia, NPs bring an opportunity to improve primary care access for rural communities. However, barriers to bringing in NPs include lack of funding for NP positions and restrictions on scope of practice (Martin-Misener et al. 2010). Rural areas also have recruitment and retention challenges for other healthcare workers including nurses. For example, Nova Scotia recently became one of the first Canadian provinces hit by a nurse shortage, with specialties including critical care nurses (Doucette 2015). Retention is an ongoing challenge in rural Nova Scotia hospitals as new nursing graduates will often work in rural areas temporarily to gain experience and then move to Halifax or out of the province. More educational opportunities along with good infrastructure, remuneration, workplace organisation, professional environment, and other support structures are all part of the retention strategy for rural health care staffing (Baernholdt and Mark 2009, Buykx et al. 2010).

What exactly is meant by "rural"? The term suggests pastoral landscapes with low population density due to isolated communities working to support industries such as farming, fishing, or logging. However, only a small fraction of rural populations are typically involved in farming with towns ranging from a handful of residents to more than ten thousand residents. Proximity of rural areas to urban centres can range from a few kilometers to hundreds of kilometers. A small rural town may only have one or two PCPs. Others may have a community hospital staffed with an Emergency Department (ED) physician but no specialty services. A larger rural town may have a regional hospital that serves patients from smaller towns providing access to specialists and surgeons. Such regional hospitals may not offer all specialty services but can facilitate the transfer of patients to urban centres when the need arises (Hart et al. 2005, DesMeules 2006).

In the US, there are at least four classification systems used to distinguish rural populations. The Office of Management and Budgets (OMB) metropolitan and nonmetropolitan

populations are county-based definitions used to determine reimbursement levels for more than 30 federal programs including Medicare. A metropolitan area is defined as a central county with one or more urbanized areas greater than or equal to 50000 residents with outlying counties that are tied to the core. With this definition the U.S. has 1090 metropolitan counties and 2052 nonmetropolitan counties. Another county-based definition that builds on the OMB dichotomy are Urban Influence Codes (UIC) such that counties are classified into nine groups: two metropolitan and seven nonmetropolitan. The groups are based on adjacency to metropolitan counties with a minimum work commuting threshold. The Census Bureau Rural and Urban taxonomy partitions urban areas into urbanized areas and urban clusters. Urbanized areas have populations of 50000 or more, urban clusters have populations ranging from 2500 to 49999 and smaller populations are rural. The Rural/Urban Commuting-Area (RUCA) Taxonomy builds on the Census Bureau taxonomy using zip-code based areas and distinguishes a small town where the majority of commuting is to a large city from a similar sized town where there is commuting primarily to other small towns. There are more than 30000 zip code areas (Hart et al. 2005). Canada uses Metropolitan Area and Census Agglomeration Influenced Zones (MIZ) to define rural and small towns. The rural definition refers to the population living outside the commuting zones of larger urban centres, specifically outside census metropolitan areas (CMAs) and census agglomerations (CAs). A strong MIZ means that 30% or more of the employed labour force lives in a CMA/CA urban core, while a moderate MIZ has at least 5%, but less than 30% of the employed labour force living in a CMA/CA urban core. A weak MIZ has more than 0% but less than 5% of the employed labour force living in a CMA/CA urban core, and no MIZ is used for communities that do not have any commuters to a CMA/CA urban core (DesMeules 2006).

In order to determine which definitions are better suited for capturing access to health care services in epidemiology studies, Hall et al. (2006) applied different definitions of rural to breast cancer incidence rates. Compared to invasive breast cancers, in situ breast cancers

typically detected through mammography, had rate ratios varying from 1.40 to 1.80 depending on which definition of rural is used. The basic finding is that dichotomous definitions may fail to capture variability in rural areas.

Rural hospitals have a patient mix with higher rates of chronic diseases (Hart et al. 2005), with more adjusted annual hospital admissions than urban residents. Average hospital length of stay (LOS) does not differ significantly between rural and urban hospitals (Reschovsky and Staiti 2005), which is expected since patient flow processes and medical procedures usually follow the same standards as those in urban centres. With fewer health care providers, rural hospitals typically do not offer the full range of therapies, and technology for some services may only be justified in urban hospitals due to cost (Ricketts 2000, DesMeules 2006). As a result, rural hospitals have a limited scope of service and are often involved in a network of health care facilities and patients are transferred to the appropriate facility when necessary (Ricketts 2000, Hart et al. 2005). Despite the high costs associated with needed health care services, U.S. Medicare payments to rural physicians and hospitals are 18% less for the same services. This is partly due to less available diagnostic services and adjustments for lower wage levels in rural communities (Ricketts 2000, Rosenthal and Fox 2000, Hart et al. 2005).

1.2 Thesis Structure and Contributions

Most healthcare operations research is focused on urban centres, possibly due to the proximity of universities to medical centres. This dissertation adopts a rural perspective which includes some of the same challenges as urban centres, but also additional challenges. In chapter 2, I begin with a detailed review of the Emergency Department (ED) Operations Management literature in order to gather a solid understanding of the state of the art concerning emergency care in rural areas. Chapters 3 and 4 are studies in collaboration with specialists in Nova Scotia, both with patient-centric models. Chapter 3 presents *Specialist Care in Rural Hospitals: From Emergency Department Consultation to Inpatient Ward Discharge*, where I study ED crowding from the perspective of Internal Medicine Specialists (Internists) with the objective to reduce patient waiting costs. This is a new perspective in

part since most ED research considers only patient flow from arrival to the ED until departure from the ED. The traditional rural situation is particularly challenging since Internists take on a dual role as the Intensive Care Unit (ICU) physician and Internist on call for ED consultation and inpatient care in the Medicine wards. In Chapter 4, *Dialysis Facility Network Design* is presented, with a rural perspective. Here the objective is to design a network of dialysis facilities that reduces long patient travel times, improving the welfare of dialysis patients from rural areas. This perspective differs from an urban-centred mindset where the number of patients needs to meet a certain threshold before service is provided. The model presented considers the reality that each patient should matter, and provides a framework to provide the best network of dialysis facilities with consideration of budget and capacity management constraints. Concluding remarks and future research is specified in Chapter 5.

The projects undertaken in this thesis address understudied research problems, and as a result, we make a significant contribution to the literature on healthcare operations research. The models presented in this thesis are of theoretical interest and also provide practical insight into solutions to address strategic and operational challenges of specialists in rural areas.

In terms of acute care operations research, Emergency Department (ED) crowding has been studied extensively for decades, yet the ED crowding problem is still a regular problem in hospitals around the world. The focus has mostly been on ED patient flow from arrival until admission or discharge from the ED, leaving little opportunity to address possibly the most difficult and imporant challenge: how should hospitals manage interdepartmental patient flow encompassing both the ED and inpatient wards? For the existing studies on ED outflow to inpatient wards, typically inpatient beds are modeled as servers and inpatient bed capacity is the targeted resource constraint. However, the most valuable resources in any hospital are the healthcare professionals who provide appropriate services for patient care. In the first project of this thesis, we consider the workflow decisions of specialists on call for acute care purposes. These decisions are interdepartmental since specialists go to

the ED for consultation and determine if patients can be discharged from the ED or require admission to inpatient wards. After admission, specialists take care of inpatients in the wards for several days until inpatient discharge. We provide a stochastic dynamic programming framework for the workflow decisions of specialists, and incorporate the additional challenges for specialists on call in rural hospitals. The problem is particularly challenging in the traditional rural situation where an Internal Medicine specialist (Internist) takes on a dual role as ICU physician and Internist on call. Our proposed framework for Specialist Care encompasses both Single Role and Dual Role models. This decision making problem could have been specified with more variables resulting in the need to use approximation methods to identify appropriate solutions. Instead, we develop models with novel elements and apply appropriate assumptions observed in practice, so that the problems can be solved by backward induction.

Chronic care is also a significant challenge for healthcare systems around the world. In the U.S., among the costs to deliver chronic care include more than $27 billion to provide dialysis services to patients with chronic kidney failure (USRDS 2015). In this thesis, the strategic problem of dialysis facility network design is studied. Our contribution involves considering the rural perspective of this problem where dialysis patients may need to travel great distances in order to receive four hours of treatment, three times per week. In addition to providing dialysis at facilities, home dialysis is also regularly provided as an option and patients have the choice of facility-based dialysis or home dialysis. This study provides the first optimization model for dialysis services planning that incorporates patient choice for dialysis mode. While cost studies have been conducted to consider the overall average cost of dialysis by modality, our optimization model includes a more detailed cost constraint, where costs related to facility location decisions and costs related to per patient treatment costs are separated accordingly. Our proposed model is a mixed integer program (MIP) that can be solved with a standard commercial MIP solver such as CPLEX. A feasible solution is always possible with the proposed model due to our novel formulation that maximizes patient welfare

as much as possible given system constraints. In the application of our proposed model through a case study, we find that with the same budget as the existing facility network, it is possible for considerable reduction of maximum and mean travel times and less variability. We also illustrate the ability to plan for improvements to the dialysis facility network by determining the best location and capacity expansion decisions if additional funding is made available.

Chapter 2
Operations Research for Emergency Care: A Review

Emergency Department (ED) crowding has been a regular and significant problem in hospitals for over twenty years (Lynn and Kellermann 1991). More than 1000 studies have been conducted by emergency medicine, computer science, and operations management researchers to investigate ED crowding. While several years of research has examined causes, effects, and proposed solutions, the problem has not yet been solved and ED crowding is getting worse (Pitts et al. 2012).

Hoot and Aronsky (2008) reviewed the medical literature on ED crowding. Commonly reported causes include increases in patient arrivals, inadequate staffing, and boarding of admitted patients. The effects are serious: adverse outcomes including patient mortality, reduction in quality of care, patient dissatisfaction and an increase in the number of patients who leave without being seen (LWBS) by a physician. When patients LWBS, the risk to patient safety is a real concern, especially among patients with severe medical conditions. While the most critical (triage level 1) patients are always given priority and treated in a separate resuscitation track, ED crowding does create delays for both low-acuity (triage level 4, 5) and high-acuity (triage level 2, 3) patients (McCarthy et al. 2009).

A common misconception is that ED crowding is due to non-emergency patients that seek care in EDs but should be treated elsewhere. When the medical community first raised the issue of ED crowding in the United States, policy makers inaccurately concluded that crowding was due to inappropriate use of emergency services by those with no urgent conditions (Olshaker and Rathlev 2006). Hospital management was left to use their existing resources and cope with ED crowding themselves. To help address this misconception, additional lobbying ensued with supporting research. Noteworthy is a large study of 110 EDs

by Schull et al. (2007) who showed that typically, low-complexity patients impact non-low-complexity patients with negligible increases in average time to see a physician and average LOS. The reality is that diverting low-complexity patients away from EDs is unlikely to improve the situation for sicker patients.

There is published evidence that ED crowding creates adverse effects on quality of care including patient mortality. The evidence includes a study of 25 community and teaching hospitals in Ontario, Canada demonstrating that ED crowding has a real impact on the time to deliver thrombolysis, with increased door-to-needle time for patients with suspected acute myocardial infarction (commonly known as a heart attack). In networks of EDs with moderate crowding, delays were observed attributable to 3 additional deaths per 1000 patients treated. Higher network crowding was shown to have more severe effects with delays attributable to 7 additional deaths per 1000 patients treated (Schull et al. 2004).

Emergency Medicine researchers have completed several empirical studies on ED crowding and its impact on quality of care. Consequences include reduced quality of pain care (Hwang et al. 2008), delays in analgesia treatment in patients with acute abdominal pain (Mills et al. 2009), higher risk of adverse cardiovascular outcomes in patients with chest pain (Pines et al. 2009), increased time to antibiotics for patients with community-acquired pneumonia (Fee et al. 2007, Pines et al. 2007), and for infants presenting with fever (Kennebeck et al. 2011). For a review of the medical literature from 1989-2007 on the effect of ED crowding on quality of care, see Bernstein et al. (2009).

Interventions to reduce average length of stay (LOS) in the ED include: physician in triage (Holroyd et al. 2007), physician orders at triage (Russ et al. 2010), bedside registration (Gorelick et al. 2005), and point-of-care (POC) testing in the ED rather than central laboratory (Jang et al. 2013). Additional interventions include fast-track facilities (Yoon 2003), increasing ICU capacity (McConnell et al. 2005) and the introduction of a computerized consultation management system (Cho et al. 2011).

We conducted a systematic review of the academic literature with the search criterion: ("Emergency Department" OR "Emergency Room" AND (Crowding OR Simulation OR Queuing OR Queueing OR Scheduling OR Staffing).

Sources used for this review include both the Institute for Scientific Information (ISI) Web of Science (Thomas Reuters) and specific databases for each of the following journals:

- Decision Analysis
- Interfaces
- Management Science
- Manufacturing and Service Operations Management (MSOM)
- Operations Research
- Production and Operations Management (POM)
- The European Journal of Operational Research (EJOR)
- Journal of the Operational Research Society (JORS)
- Health Care Management Science
- Medical Decision Making
- Annals of Emergency Medicine
- Academic Emergency Medicine

The ISI Web of Science search was confined to the following four Research Areas:

- EMERGENCY MEDICINE (EM)
- HEALTH CARE SCIENCES & SERVICES (HCSS)
- COMPUTER SCIENCE (CS)
- OPERATIONS RESEARCH MANAGEMENT SCIENCE (ORMS)

Using this criterion, ISI identified a total of 1368 journal articles. Of those, 1095, 213, 48 and 29 were from EM, HCSS, CS, and ORMS respectively. (The CS results include 10 papers that are also included in HCSS and 7 papers that are also included in ORMS).

Journal-specific searches returned a total of 484 papers, the majority coming from Academic Emergency Medicine (226) and Annals of Emergency Medicine (148).

Abstracts of the articles were reviewed to identify papers with a specific focus on operations research and management science modeling of Emergency Department (ED) processes. Our key interest is to identify studies that pay particular attention to the rural context.

2.1 Measuring ED Crowding

While ED crowding is a widely reported problem, there is no standard way to measure it (Hwang and Concato 2004). A group of 74 experts developed 113 potential measures, reduced to a final list of 38 different measures of ED crowding (Solberg et al. 2003). Measures include occupancy levels, waiting times, LWBS, and length of stay (LOS). A variety of indices have been proposed to measure ED crowding including the National Emergency Department Overcrowding Score (NEDOCS) (Weiss et al. 2004), the Emergency Department Work Index (EDWIN) (Bernstein et al. 2003), the Demand Value of the Real-time Emergency Analysis of Demand Indicators (READI) (Reeder et al. 2003), and the Work Score (Epstein and Tian 2006).

NEDOCS is scaled from 0 to 200 with 100 set as the cutoff for overcrowding. There are six levels separated as follows: 0 - 20 = Not Busy, 21 - 60 = Busy, 61 - 100 = Very Busy, 101 - 140 = Overcrowded, 141 - 180 = Dangerous, Above 180 = Disaster. Scores can be calculated using the following equation:

$$\begin{aligned} NEDOCS_t &= -20 + 85.8 * (TotalPatients/EDBeds) \\ &\quad +600 * (Admits/HospitalBeds) + 13.4 * (Ventilators) \\ &\quad +.93 * (LongestAdmit) + 5.64 * (LastBedTime) \end{aligned}$$

where:

Total Patients	=	number of total patients in the ED at the time the score is calculated
ED Beds	=	total number of ED beds including hallways, chairs, fast track and other beds that can be used to serve patients at the time the score is calculated
Admits	=	number of holdovers/admits in the ED, at the time the score is calculated
Hospital Beds	=	Total number of hospital beds
Ventilators	=	number of patients on ventilators/respirators in the ED at the time the score is calculated
Longest Admit	=	longest admit holdover/boarding (in hours) at the time the score is calculated
Last Bed Time	=	wait time (in hours) from arrival to bed for the last patient called for a bed

Using data from an 8-week period from Vanderbilt University Medical Center, Hoot et al. (2007) conducted a study to test the usefulness of EDWIN, NEDOCS, READI, and the Work Score to measure present and predict future ED crowding. A program was developed in Matlab to query the information system every 10 minutes, extracting occupancy level and other data needed to calculate the 4 crowding measures. Ambulance diversion was used as the outcome measure for ED crowding and receiver operating characteristic (ROC) curves were plotted for the analysis. The area under the ROC curves were 0.81 for EDWIN, 0.88 for NEDOCS, 0.65 for READI, 0.90 for the Work Score, and 0.90 for occupancy level. The authors conclude that the simplest measure, ED occupancy, performs as well as the more complicated indices. However, the ambulance diversion policy adopted at the hospital under study is to go on diversion whenever any of the following apply and are not remedied within an hour:

- All critical care beds in the ED are occupied, patients are occupying hallway spaces, and at least 10 patients are waiting
- An acuity level exists that places additional patients at risk
- All monitored beds within the ED are full

Obviously, ED occupancy is a good predictor of ambulance diversion when the trigger for ambulance diversion is based primarily on ED occupancy. It may be true that simpler measures are better suited than more complicated indices to measure ED crowding, but we still do not have a standard way to measure it.

2.2 ED Patient Flow Process

EDs operate in a complex multi-server environment with time-varying patient arrivals, abandonments, multiple patient classes with priority queues, and a multi-stage treatment/service process. ED care can be conceived to have three main components: *waiting room time* from arrival up to the time when a patient is placed in an ED bed or other treatment area, *treatment time* from ED bed placement up to the admit/discharge disposition decision and *boarding time* for admitted patients which runs from disposition decision until the patient is transferred to a specialty ward for additional care. Figure 2-1 represents the ED patient flow process.

Patients may walk-in or arrive to the ED by ambulance. Arriving patients are clinically assessed initially in triage, typically by a triage nurse who assesses the patient's medical condition to identify a priority for treatment. Following triage, registration is completed to identify and record patient information. Due to long waits or other factors, some patients will leave without being seen (LWBS) by a physician and exit the ED. Patients wait in the waiting area until they are assigned an ED bed or other area for treatment. Note that the waiting area is a priority queue since higher risk patients are served first. ED treatment involves several components, typically beginning with both ED physician and nurse assessments. After this initial diagnosis, an ED physician may investigate by requesting one or several lab tests (i.e. blood/fluids or imaging), keep the patient under observation, and may request a consultation from a physician specialist. Consultations may be for assessment purposes or for admission approval. Consultations made for assessment purposes may include a clinical examination of the patient by the specialist and additional lab tests (further investigation). In order to admit a patient to stay in the hospital in another department, a specialist may

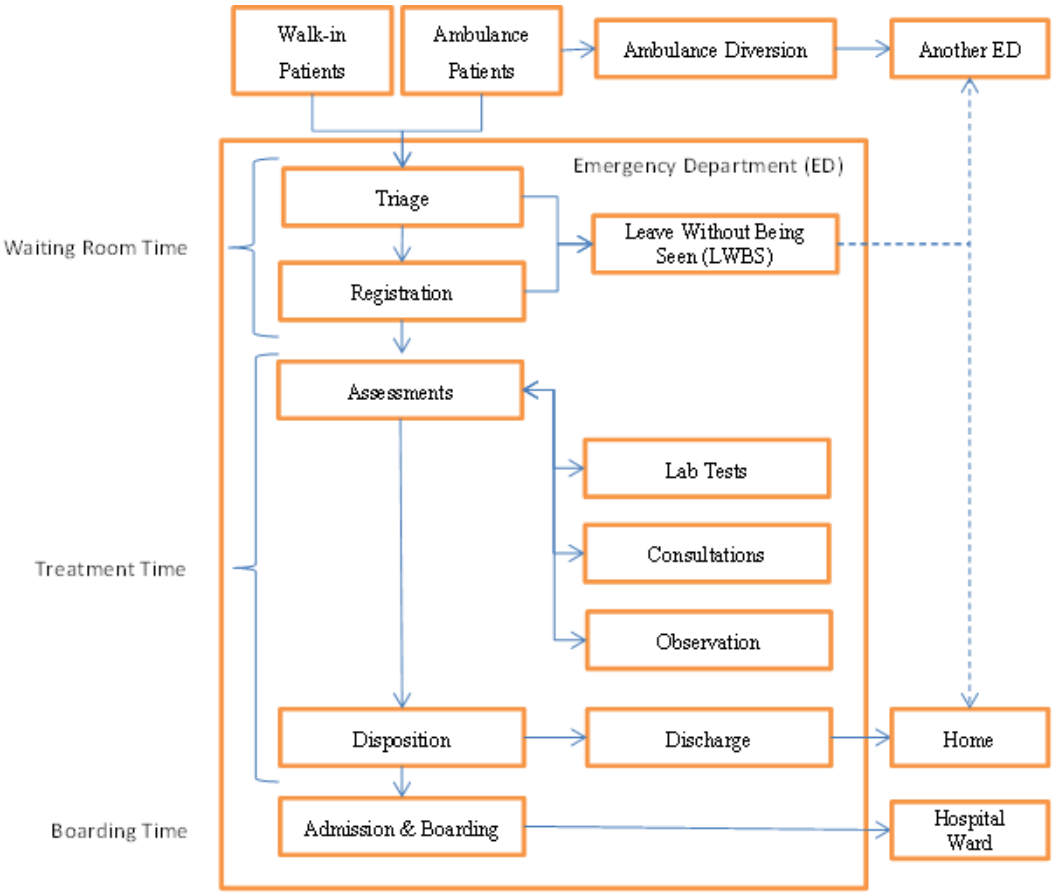

Figure 2–1: ED Patient Flow

need to be consulted from an admitting department (inpatient ward). The last step in the ED treatment process is the disposition decision to either admit the patient to the hospital for further treatment or to discharge the patient home. A discharge process or admission and boarding process follows before the patient exits the ED. While admitted patients are waiting in the ED for an inpatient bed, they are often moved out of the ED bed but remain boarding in hallways.

2.3 Literature Review

Emergency medicine researchers have been interested in the application of operations management (OM) techniques to help address the ED crowding issue for more than ten years. To encourage researchers to examine the problem from a systems perspective, Asplin et al. (2003) proposed a framework for ED crowding research based on the *input-throughput-output* model.

The input component includes three categories of demand for ED care: 1) *emergency care*, 2) *unscheduled urgent care*, and 3) *safety net care*. *Emergency care* is provided in the ED for seriously ill and injured patients. It also includes referrals from other health care providers who anticipate the need for patient admission to a hospital. *Unscheduled urgent care* arises due to inadequate capacity in other parts of the healthcare system. The delay for an acute care appointment may be longer than patients are willing or able to wait. Considering that many clinics are only open during day time hours whereas EDs are always open to the public, 24 hours per day, some patients prefer to wait in the ED and receive same-day care. *Safety net care* is provided in the ED to patients who have limited or no other place to go for medical care. Access barriers are particularly prevalent in countries such as the United States (US) where a large number of uninsured citizens can only seek care in EDs. However, even in Canada's publicly funded health care system, where health care services are provided universally to all citizens, access issues exist due to other problems. For example, in the province of Quebec, many citizens do not have access to a primary care physician, so they seek care in EDs instead.

The throughput component includes internal ED processes, proposed as two phases. The first phase includes triage, room placement, and initial evaluation. The second phase includes diagnostic testing and ED treatment.

The output component focuses on the patient disposition decision to either discharge home or admit to a hospital ward for further care. From an ED patient flow perspective, the main bottleneck is the inpatient exit rate, so inpatient boarding of admitted patients is considered a particularly important research area. For discharged patients, unscheduled return ED visits may occur due to inappropriate discharge or inadequate access to follow-up care.

Soremekun et al. (2011) illustrate the concept of the ED efficient frontier. They explain how the efficient frontier provides a way to compare multiple EDs in terms of responsiveness (1/wait time) and utilization rates. In line with standard OM practice, three ways are suggested to move towards the efficiency frontier: 1) eliminate waste, 2) reduce variability and 3) increase flexibility.

Eliminating waste is the key principle of lean management, where process improvement teams identify non-value added activities of each process from the patient's perspective. Several examples of successful Lean projects are included in a recent review of lean implementations from 15 EDs in Australia, Canada, and the United States (Holden 2011). Examples include both eliminating outdated policies and developing new concepts such as fast-track for low-complexity patients, patient streaming according to probability of admit/discharge, streaming into 3 "pods" (complex, medium, and fast), and a new process for pulling patients into inpatient wards.

Reducing variability includes both demand and service time variability. Considering the safety net role of the ED, there are few options to reduce demand variability, although interventions such as ambulance diversion may result in some reduction of ED demand. However, several interventions exist for reducing ED service time variability. Reducing variation in physician and other provider practices will clearly lead to lower service time variability.

Other examples include decreasing test utilization and/or response times and smoothing surgical schedules.

Increasing flexibility is an important option, especially considering that a significant amount of ED demand is predictable. Staffing based on models that incorporate time-varying demand creates the flexibility required to better match demand patterns. Having an on-call system for additional staff provides reactive capacity to handle unpredictable demand. Canceling elective surgeries is another example that can reduce ED boarding by providing reactive inpatient bed capacity.

Most of the remaining papers in the literature review are organized according to ED patient flow in the following sections: 1) Demand for ED Services, 2) Ambulance Diversion and Offload Delays, 3) Triage, 4) Care and Treatment, 5) Admission & Boarding. We then review methodological alternatives.

2.4 Demand for ED Services

The demand for ED services can be considered either from the perspective of one specific hospital or from the perspective of a government (or other organization) responsible for multiple hospitals within a geographic area. Most of the studies in this review consider problems from the perspective of ED demand for a specific hospital; however, in some cases, major system design changes are also conceived. For example, Congdon (2001) investigates potential system changes such as the impact of expansion and closure of Accident & Emergency (A&E) departments in England. Patient flows are modeled using gravity models to match demand from patient populations with the supply of A&E facilities in different hospitals. Both distance and travel time based accessibility models are considered, as well as an extended distance-based model that incorporates additional factors such as a regional variation in health need. Facility network design models that incorporate congestion have also been developed by Zhang et al. (2010) in the context of preventive healthcare.

From the perspective of a given hospital, historical data shows that daily ED census is cyclical, with predictable patterns depending on time of day, day of week, and time of

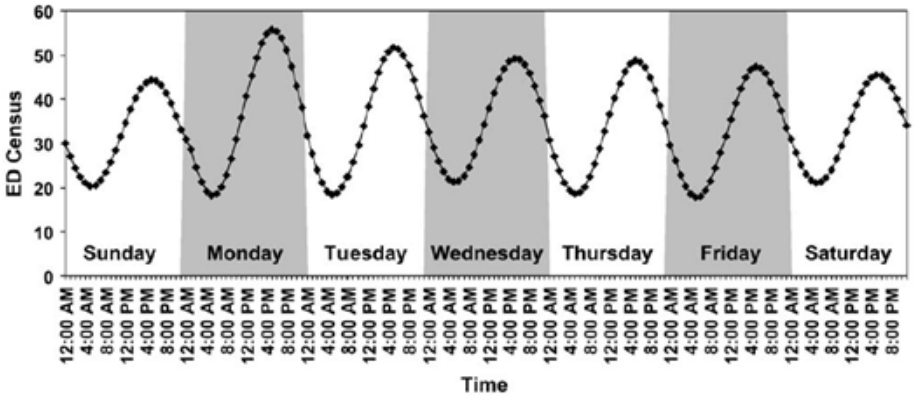

Figure 2–2: ED Census

the year. These patterns are based on time-varying patient arrival rates. For example, in a Level 1 trauma center in St. Paul, MN, ED census peaks occur in the afternoon and remain high in the evening. ED census then typically declines until early the next morning when it begins to gradually increase again. While the trend is similar the next day, the amplitude is different depending on the day of the week, with Mondays having the highest peak. The exact timing of peak periods may be different in other EDs, but the patterns shown in Figure 2-2 are typical (Asplin et al. 2006).

Green et al. (2007) examine queueing models for analyzing service systems with time-varying demand. The focus of the paper is on call centers but includes other application areas including EDs. Their discussion is centered around the M(t)/GI/s(t) + GI queueing model, which has a nonhomogeneous Poisson arrival process with a time-varying arrival function $\lambda(t)$, independent and identically distributed (i.i.d) service times with a general probability distribution, time-varying number of servers s(t), with i.i.d times to abandonment following a general probability distribution. Note that the M(t)/GI/s(t)+GI model also assumes an infinite waiting area and a first-come first-served (FCFS) queueing discipline.

Classical approaches for dealing with time-varying demand include Pointwise Stationary Approximation (PSA) and Simple Peak Hour Approximation (SPHA). PSA is an effective analytical strategy for dealing with time-varying demand in settings with short service times (e.g. 3 minutes), a high quality of service standard and short staffing intervals. SPHA

is appropriate again when services times are short and quality of service is high, but the staffing interval is long. PSA does not deal with the fact that in practice, staffing levels generally need to be held constant during each staffing interval. Therefore, an adjusted approach called segmented-PSA or another similar approach called stationary independent period-by-period (SIPP) is often used. When services times are moderate (e.g. 30 minutes), lagged refinements to SIPP and SPHA, called Lag SIPP or Lag SPHA are recommended. In cases when service times are long (e.g. 300 minutes), then the modified-offered-load (MOL) approximation should be used instead. The study concludes that much more work is needed to examine EDs and other more complex service systems.

For an urban ED with an annual census of 25,000 patients, Green et al. (2006b) developed a queueing model to estimate the number of healthcare providers required in each shift. The study used an M/M/s model with a single queue to represent the waiting area and multiple servers to represent healthcare providers. The model does consider time-varying demands to account for the fluctuation of arrivals over the course of the day. This was done using Lag SIPP. As noted earlier, SIPP essentially constructs separate queueing systems for each staffing period. However, considering that peak congestion often lags the peak arrival rate, the Lag SIPP method was used to provide a better estimate than SIPP. In addition to variation in demand by time of the day, the overall average volume also varied for each day of the week. In this case, creating different schedules for each day of the week was considered impractical, but two separate queueing analyses were performed to identify a weekday schedule and a weekend schedule.

While the study was successful in reducing LWBS, the authors noted several limitations. The model did not incorporate priorities based on the triage system or account for any additional registration and triage delays. It also assumes service time provided by a healthcare provider is exponentially distributed and continuous; however patients often see a physician, then wait in an ED bed for lab test results or consultations and then see the physician again before admission or discharge. In addition, the authors noted that a more detailed analysis

is required, especially in larger EDs, which often include fast track areas and different types of providers such as resident physicians and nurse practitioners.

2.5 Ambulance Diversion and Offload Delays

In many regions, EDs have the option to go on *ambulance diversion*. During periods of peak congestion, ED management will declare diversion status, requesting the emergency medical services (EMS) agency to divert incoming ambulances to another hospital. In essence, ambulance diversion is an intervention to reduce ED crowding. When ambulances are diverted from overcrowded EDs to other EDs with more available capacity, resource pooling benefits should occur and result in a reduction in ED crowding. However, a lack of pooling benefits exists in practice, possibly since diversion requests may be ignored by EMS when all EDs are on diversion, commonly referred to as "All on Diversion, Nobody on Diversion" (ADND).

Deo and Gurvich (2011) studied ambulance diversion in a theoretical network of two EDs to examine why resource pooling benefits may not always be realized. Their model embeds a queueing network within a static non-cooperative game between the two EDs, each with the objective of minimizing their own waiting time. They suggest that a defensive equilibrium exists such that both EDs do not accept diverted ambulances from the other ED. The authors propose that pooling benefits could be realized if centralized diversion decisions were coordinated by a social planner rather than decentralized decisions made by each ED.

For each ED (i=1, 2), ambulance (a) and walk-in (w) patient arrivals are modeled according to Poisson processes, with rates λ_a^i and λ_w^i respectively. The EDs have N_i beds and service times are assumed to be exponentially distributed with the same mean in both EDs. Diversion status is declared at time t, if the number of patients in the ED exceeds a diversion threshold, K_i. In the decentralized situation, the Nash equilibrium of the model is the threshold pair K=(0,0), which implies that the best response for each ED is to always be on diversion, resulting in no diversion and no pooling benefits (i.e two independent M/M/N queues). For the centralized situation, it is difficult to find an analytical solution for the social

planner's optimum threshold pair, $K = (K_1^*, K_2^*)$. Instead, a lower bound of the solution is solved for a perfectly pooled system and an upper bound of the solution is solved with a capacity-based static threshold. The lower bound solution is unrealistic for the ED setting since it allows for: 1) both rerouting of ambulance and walk-in patients and 2) rerouting of patients after they have already been waiting in the queue. The upper bound solution uses the number of ED beds as a threshold pair, $K = (N_1, N_2)$. The authors propose this solution to be implemented with a policy that EDs cannot divert ambulances if there are available ED beds.

Ambulance offload delays occur when paramedics cannot immediately transfer patient care to ED staff upon arrival. When an ED is too crowded, paramedics continue to provide patient care either in the ambulance or on an ED stretcher. An empirical study by Eckstein and Chan (2004) found that ambulance offload delays occurred in 12.5% of all transports. Almehdawe et al. (2013) developed a stochastic queueing network model of the EMS-ED interface to assess the impact of system resources on ambulance offload delays.

The authors consider a multiple server queueing network model with two customer classes: ambulance patients and walk-in patients, with priority to ambulance patients. In the model, ambulance patients arrive to the system according to a Poisson process at rate λ_0. Patients are routed to one of K EDs according to a routing probability $p_k, k = 1, ..., K$. Walk-in patients are incorporated in separate arrival streams according to a Poisson process at rate λ_k for each respective ED, $k = 1, , K$. All patients in the model are assumed to require ED beds, so lower acuity patients are excluded with the assumption that they receive care in a separate "minor treatment" area. Service assumptions include exponentially distributed service times with rate μ_k, patients served on a first-come-first-served (FCFS) basis within each priority class (ambulance or walk-in) with preemptive priority to ambulance patients over walk-in patients. Ambulance transportation time is assumed negligible and patients are considered lost if all N ambulances are occupied.

The state of the system is represented by the following two variables at time t:
$q_{a,k}(t)$: number of ambulance patients in service or waiting in the kth ED
$q_{w,k}(t)$: number of walk-in patients in service or waiting in the kth ED

However, due to the assumption that ambulance patients receive preemptive service over walk-in patients, the authors analyze ambulance patients separately from walk-in patients. As a result, the system is analyzed as a continuous time Markov chain (CTMC), quasi-birth-and-death (QBD) stochastic process $(q_{a,K}(t), q_{a,K-1}(t), ..., q_{a,1}(t)), t \geq 0$ with a *finite* number of levels. Steady-state distributions are solved using matrix-analytic methods. Performance measures include the mean number of ambulances in offload delay at the Kth ED and the probability that all ambulances are in offload delay, referred to as the loss probability, P_L.

The CTMC is enhanced by adding the walk-in patient queue of one ED, obtaining a QBD process $(q_{w,K}(t), q_{a,K}(t), q_{a,K-1}(t), ..., q_{a,1}(t)), t \geq 0$ with an *infinite* number of levels. The model is tested with three case studies: 1) a small network with infrequent offload delays, 2) larger network with significant offload delays, and 3) the real EMS-ED network that motivated the study. In the small network, prioritizing ambulance patients result in a reduction of offload delays and shorter waiting times for ambulance patients, at the cost of longer walk-in patient wait times. In the larger network, the authors test the impact of varying routing probabilities. A balanced scenario where routing probabilities are proportional to ED capacity results in a 14% decrease in offload delays. In the third case study, the authors investigate the impact of changing LOS by changing service rates. As expected, offload delays, expected queue lengths and walk-in patient average LOS are all reduced.

Considering that the model has numerous assumptions, the authors also created a simulation model to relax two main assumptions: 1) ambulance travel times are negligible, and 2) ED service times are exponentially distributed. Ambulance travel times from the study setting follow a beta distribution. The simulation model tested both beta distributed travel times as well as exponential distributed travel times, due to previous models of ambulance

travel times. Service times were modeled based on total flow time, with an assumption that service time has a similar distribution to total flow time with a different mean. As a result, the simulated service times follow an Erlang distribution, and are modeled with the pre-emptive resume service discipline. If the loss probability is small, the simulated results are not significantly affected by adding ambulance travel times and changing the service time distribution. The authors conclude that considering that a typical ambulance utilization rate is approximately 35%, the loss probability is small in practice, and the model is appropriate under normal EMS operating conditions.

2.6 Triage

ED triage is the process that assesses the severity of patients' medical conditions and assigns a priority for treatment. A triage code is established for each patient upon arrival, typically by a triage nurse based on the severity of patient symptoms. The Emergency Service Index (ESI) is a 5-level triage level rating system used in the United States (Gilboy et al. 2005). In terms of validity and reliability, 5-level triage systems are better at assessing patient severity than three-level triage systems. As a result, more patients are now triaged using ESI than any other triage system in the United States (McHugh et al. 2012). Other triage systems include the Australasian Triage Scale (ATS) which, according to FitzGerald et al. (2010), formed the basis for the Manchester Triage Scale (MTS) in the UK as well as the Canadian Emergency Department Triage and Acuity Scale (CTAS) in Canada. CTAS levels are shown in Table 2-1. ESI uses the same levels as CTAS except level V is named Referred instead of Non Urgent. CTAS includes national guidelines on the maximum time patients should wait until being seen by provider for each acuity level. The guidelines also include fractile response objectives for the proportion of patients that should be seen within the time frame for each level.

According to the Canadian Association of Emergency Physicians (Beveridge et al. 1999), triage goals are:

1. To rapidly identify patients with urgent, life threatening conditions

Table 2–1: CTAS levels and guidelines

CTAS	Time to physician	Fractile response objective
Level I Resuscitation	Immediate	98%
Level II - Emergent	15 minutes	95%
Level III Urgent	30 minutes	90%
Level IV Less Urgent	60 minutes	85%
Level V Non Urgent	120 minutes	80%

2. To determine the most appropriate treatment area for patients presenting to the ED
3. To decrease congestion in emergency treatment areas
4. To provide ongoing assessment of patients
5. To provide information to patients and families regarding services, expected care and waiting times
6. To contribute information that helps to define departmental acuity

Triage requires quick decision-making with limited information, which naturally results in inaccurate triage codes being assigned to some patients. In a study focused on triage of children presenting to a pediatric ED with abdominal pain, Wilk et al. (2005) developed an approach to help improve triage accuracy. Abdominal pain is the main symptom for patients with acute appendicitis, a serious medical condition with high mortality rates if not treated promptly. Only a very small number of children with abdominal pain have appendicitis though, and unnecessary investigations and assessments are costly to hospitals and painful for patients.

Patients who arrive to the ED with abdominal pain are examined by a triage nurse practitioner (NP), followed by an evaluation by an ED physician. If acute appendicitis is suspected, a surgeon is called in for consultation. Otherwise, the evaluation is either not-yet-diagnosed (NYD) or resolved. Resolution occurs if the child's abdominal pain subsides naturally and the patient is discharged. These three evaluations: *surgical consult*, *NYD*, and *resolution*, correspond to the three main record classes in the study data set of 647 patient records, each with 12 attributes.

The proposed methodology is based on rough set theory, fuzzy measures, and game theory. Rough sets use information tables, where rows represent objects such as patient charts, and columns represent attributes. These attributes include both condition attributes and decision attributes. In the studied case, there are 12 condition attributes and one decision attribute, referred to as the *triage outcome*: surgical consult, NYD, or resolution. Using knowledge expressed in the condition attributes, an approximation of knowledge is established for the decision attribute by the rough set.

How to handle missing values within the data set is stressed as an important issue. Records with missing values may be quite important and should not be discarded. For example, the attribute white blood cell count (WBCC) is missing in 44% of the records with resolution class, but missing in only 3% of the surgical consult class. WBCC is not an important attribute for diagnosis of patients in the resolution class, but it is quite important in the surgical consult class. So the attribute may be important depending on the context, and discarding all records that are missing WBCC would not be appropriate. Traditional rough set theory does not handle missing values so the rough set is extended to include missing values.

Fuzzy measures and game theory are used to model the relative value of the information supplied by each attribute. For the set C of all condition attributes, which represent the players in a game, the characteristic function $\mu(A)$, represents the payoff obtained from a cooperative game from a coalition of attributes. Shapley values (Shapley 1952) are often used to calculate solutions for cooperative games, interpreted in this context as the average contribution of an attribute to all possible coalitions of attributes from C. Therefore, attributes with higher Shapley values are considered to be the most important attributes that best explain relationships within the data set. Decision rules are then generated using the Explore algorithm. The NYD class could not be predicted with sufficient accuracy. However, promising results from the other two classes has led to the development of a decision support system and further research on improving triage accuracy.

Opportunities to improve operational efficiency naturally exist at triage, where appropriate information may be utilized to create separate patient streams. For example, many EDs have separate *fast-track* facilities designed to serve low-acuity patients whose expected treatment time is shortest (ESI/CTAS levels 4 and 5). Such facilities improve patient throughput, decrease costs and increase patient satisfaction (Yoon 2003).

In order to compare the fast-track triage approach with an alternative acuity ratio triage (ART) approach, Connelly and Bair (2004) developed a discrete event simulation model for a Level 1 trauma center at the University of California, Davis, Medical Center (UCDMC). Rather than having a separate fast-track facility for low-acuity patients, the alternate ART approach involves assigning a ratio of high-acuity (HA) to low-acuity (LA) patients to each health care provider, without a separate fast-track area.

To model the fast-track triage approach, HA and LA patients are treated in completely separate areas, and if HA/LA healthcare staff members have downtime, they do not cross over to serve LA/HA patients in the other area. However, imaging and lab facilities are shared by the HA and FT areas. On the other hand, the model of the ART approach used a single mixed-acuity treatment area that managed all patients (59% HA and 41% LA). Otherwise, all other simulated parameters, including patient population, staff, and bed capacity, were the same for the two scenarios.

Patient inter-arrival times follow an exponential distribution and staff activities are prioritized in job queues according to patient acuity. In emergency cases for trauma and resuscitation, preemptive service is provided over lower-priority cases. The model includes activities for imaging, lab tests, history and physical examination, consultations, and other procedures. Model accuracy is within 10% for average times, however, individual patient times are far from accurate with errors of more than 3 hours in more than 50% of cases.

Tested scenarios include comparing the ART and FT triage systems with equal HA:LA ratio as well as a ratio of 12:8 to represent that actual ratio used in practice during the study period. The results showed reduction in average wait times for HA patients with ART

at the cost of much higher average wait times for LA patients. The authors conclude that definitive conclusions cannot be drawn from their results and further research is needed to improve the predictive capability for individual patient times.

As part of a system-wide redesign effort by Banner Health (BH) in Arizona, Cochran and Roche (2009) developed a queueing network model of the ED. In their split patient flow (SPF) model, high-acuity and low acuity patients are "split" into two separate streams. High acuity patients are treated in traditional ED beds, while lower-acuity patients are fast-tracked into a separate stream where they wait in chairs and walk between treatment areas for physician assessment, procedures, and wait for test results, if required. The authors analyzed seven BH EDs and report significant reductions in door-to-doc (D2D) times and the proportion of patients who leave without treatment (LWOT).

The model incorporates 5-levels of patient acuity and has nine nodes in the queueing network, the first being Registration (R) which includes triage. In the model, high acuity patients correspond to triage level 1 and 2 patients, while low acuity patients correspond to triage level 3, 4, and 5 patients. After triage, high acuity patients generally use traditional "in-patient" ED services, denoted IP_{ED}, modeled with the following nodes in the queueing network: IP_{ED} (E), Observation (O), Behavioral Health (B) and Admit Hold (A). On the other hand, low acuity patients use "out-patient" ED services, denoted OP_{ED} and are routed to a separate set of nodes: Intake (I), OP_{ED} Discharge (D), Results Waiting (W), and Procedures (P).

Node capacity is generally based on the number of beds (or chairs in areas such as Results Waiting), with the exception of the Intake area of OP_{ED}, where capacity is based instead on the number of physicians. Most routing between nodes is completely separate with high acuity patients following IP_{ED} nodes and low acuity patients following OP_{ED} nodes. However, while all level 4 and 5 patients only use OP_{ED} nodes since they are assumed to be fast-tracked and discharged, a fraction of level 3 patients may move from OP_{ED} nodes to

IP_{ED} nodes. In the case of routing level 3 patients from Results Waiting to the IP_{ED}, this fraction is approximated based on admission rates of level 1 and level 2 patients.

Routing matrices are developed for each of the five patient acuity levels, and the probability of flow from node i to node j for patient acuity type t, is denoted r_{ij}^t. The SPF ED model only has forward flow through the queueing network. The arrival rate to each node i is calculated accordingly using:

$$\lambda_i = \sum_{t=1}^{5} \gamma_i^{(t)} + \sum_{t=1}^{5} \sum_{j=1}^{9} \lambda_j^{(t)} r_{ji}^{(t)}$$

where: $\gamma_i^{(t)}$ = the external arrival rate for patient t (0 at all nodes except Registration), $\sum_{j=1}^{9} \lambda_j^{(t)} * r_{ji}^{(t)}$ = the arrival rate of patient type t transferred from all other nodes to node i.

Time-varying ED arrivals are also noted, and the BH study EDs have a 12 hour peak period from 9am to 9pm. The authors suggest that yearly ED volume can be converted to hourly patient arrivals using a seasonality multiplier and a peaking multiplier. Performance measures such as patient wait times (W_q) are calculated using an M/G/c approximation, and overflow probabilities (p_c) are calculated using an M/G/c/c approximation. D2D times are then calculated using the following formula:

$$\begin{aligned} D2Dtime \;=\; & W_{q_{REG}} + LOS_{REG} \\ & + \frac{f^{(1)} + f^{(2)}}{f^{(1)} + f^{(2)} + f^{(3)} + f^{(4)} + f^{(5)}} * (traveltime_{IPED} + W_{q_{IPED}}) \\ & + \frac{f^{(2)} + f^{(3)} + f^{(4)}}{f^{(1)} + f^{(2)} + f^{(3)} + f^{(4)} + f^{(5)}} * (traveltime_{INTAKE} + W_{q_{INTAKE}}) \end{aligned}$$

where: $f^{(t)}$ = fraction of patient type t

Travel times are estimated by hospital staff. Other data used in the SPF model is drawn from a variety of sources ranging from actual data to hospital staff estimates. In cases when service time distributions cannot be drawn from actual data, exponential distributions are used. The model assumes $LWBS = 0$, with a single patient class, using approximate

performance measures for wait times (e.g. W_q) and assumes there is no blocking between areas.

Patient streaming has also been proposed to partition patients into two other streams: one for "D" patients most likely to be discharged from the ED and another for "A" patients most likely to be admitted to the hospital (Saghafian et al. 2012). Considering that resource pooling generally leads to improved resource utilization and better operational efficiency, the benefits of a patient streaming initiative must outweigh the anti-pooling disadvantage. Motivated by studies in Australia where triage nurses were able to predict the admit/discharge disposition decision with roughly 80% accuracy (Holdgate et al. 2007, King et al. 2006), analytic and simulation models were used to investigate if such a streaming policy can help improve ED performance.

ED patient flow has multiple phases and the authors refer to two phases of sequencing decisions. Phase 1 sequencing determines the order that patients are taken from the waiting area to the examination room. Phase 2 sequencing decisions are made by physicians who decide the order in which patients are seen (which can be based on time in system, ESI level and other patient factors).

Considering that ESI 1 patients have serious medical conditions and are already segmented in a separate *resuscitation track*, and ESI 4 and 5 patients are often already segmented in a separate *fast-track* facility, the study focuses on ESI 2 and 3 patients. Two metrics are used in the study: length of stay (LOS), measured as the total time in ED from arrival to discharge/admit, and time to first treatment (TTFT), measured as the time from arrival to the first physician assessment. The authors argue that LOS is the key metric for D patients but TTFT is the key metric for A patients. The analytical models examine policies, denoted π =PA(Pooling with priority to As), PD (Pooling with priority to Ds), S(Streaming), and use an objective function: $\beta TTFT(A) + (1-\beta)LOS(D)$, where β = relative weight placed on TTFT of A patients.

The first analytical model is a *clearing* queueing model with a number of simplifying assumptions for tractability, including that all patients are available at the beginning of the day, with only two physicians (one for the A stream and one for the D stream), perfect A/D classification, and patient diagnosis/treatment as a single stage service (phase 1 sequencing only).

A multistage analytical model is also examined to study the effect of Phase 2 sequencing. To make this model tractable, additional simplifications include considering only one patient class (a single ESI level), along with assumptions that: there are enough examination rooms to hold all patients, all services times in wait and treatment states are i.i.d. (independent and identically distributed) and exponentially distributed (without any queueing for lab tests or other services) and preemptive service is allowed. As one might expect with this model, D stream physicians should adopt a Prioritize Old (PO) policy (since LOS matters most for D patients) while A stream physicians should adopt a Prioritize New (PN) policy (since TTFT matters most for A patients).

The simulation model includes multiple customer classes with time-varying arrivals according to nonstationary Poisson processes, and a multistage service process with several phases (up to 7) of patient-physician interactions/treatment followed by tests and preparations. The number of interactions is simulated based on data from another study (Graff et al. 1993). The service process is noncollaborative (ED physician generally does not transfer patients to another ED physician) and nonpreemptive (ED physician generally does not move to another patient in the middle of the current interaction). In the simulation model, different Protocol/Phase 1/Phase 2 scenarios were considered with:

Protocol	Pooling (P), Streaming (S) or Virtual Streaming (VS)
Phase 1	ESI or (AD + ESI)
Phase 2	Service in Random Order (SIRO) or First Come First Served (FCFS) or Prioritize New Prioritize Old (PNPO)

Note that the difference between S and VS is that resources are physically segregated and cannot be shared in S whereas VS is logically segregated with the ability to share resources across streams. Antipooling effects are so significant in S that streaming is only an attractive option if implemented as VS. Also note that PNPO represents Phase 2 sequencing in which A stream physicians prioritize new patients and D stream physicians prioritize old patients.

The results from the simulation analysis found that the VS/AD+ESI/PNPO patient flow design is an attractive option that can improve ED performance. The authors conclude that virtual streaming will be most effective in an ED with 1) a high percentage of As, 2) longer service times for As than Ds, 3) long patient boarding times, 4) high day-to-day variation in patient mix, and 5) high average physician utilization.

2.7 Care and Treatment

Early ED research includes stochastic modeling of ED patient flows. Panayiotopoulos and Vassilacopoulos (1984) developed a stochastic simulation model of ED patient flow for hospitals in Greece. They developed a simulation of a $(GI/G/m(t)) : (IHFF/N/\infty)$ queueing model with a general independent (GI) arrival distribution, general (G) service time distribution, variable number of servers, $m(t)$, IHFF queueing discipline, Finite capacity, N, Infinite customer population, ∞, single arrivals, and no unserved customers leave the system (implies LWBS=0). Panayiotopoulos and Vassilacopoulos use a variation of Lee's extension (Lee 1966) to Kendall's notation (Kendall 1953) for queueing systems. The standard A/B/C/D/E/F notation denotes A as the input process, B as the service mechanism,

C as the number of servers, D as system capacity, E as the size of the customer population, and F as the queue discipline. In this case, the standard queueing notation would be $GI/G/m(t)/N/\infty/IHFF$.

The IHFF queueing discipline adopted by the authors is an abbreviation of INCRP/HOL/FINCRP/FCFS whereby patients are prioritized according to INCReasing-Priority (INCRP) numbers with the Faster-INCReasing-Priority numbers (FINCRP) among the Head-Of-Line (HOL) patients being served first. If there are any ties in priority numbers, then those patients are served on a First-Come-First-Served (FCSFS) basis. HOL refers to the patient in the queue with the highest priority number and the queueing discipline adopted is non-preemptive.

The simulation model represents an ED where service is offered 24 hours per day, with a finite number of physicians who are each assumed to be present in the ED for one shift per 24 hours. Coded in Fortran 77, the program simulates a 24-hour period of ED activity in approximately three minutes of run time. Interestingly, while the model incorporated arrival and service events which are also included in any recent stochastic queueing system, the model also includes events corresponding to *changes of priority numbers*, which are not included in recent models. According to the authors, this was incorporated since, if a patient's condition worsens while waiting, a new higher priority number would be assigned.

Numerical examples are provided in the paper to show how different physician staffing policies result in a simulated reduction in metrics such as average time in the system and average time in the queue. The authors state that two Greek hospitals achieved a 30% improvement after adopting their methodology and other hospitals were also interested in applying the same technique.

In another early study, Vassilacopoulos (1985) uses dynamic programming models to allocate doctors to shifts in an accident and emergency (A&E) department in the United Kingdom (UK). Considering time-varying patient arrivals by hour of day and day of week, a constant patient arrival rate (λ_t) for each hour (t) of the week is considered. With the

assumption that the number of doctors (λ_t) can only be changed on the hour, a discrete problem is modeled. The real number (rather than integer number) of doctors to allocate per hour of week (q_t) is calculated to be proportional to the arrival rate using the equation:

$$q_t = \frac{M}{\lambda}\lambda_t, t = 1, 2, ..., 168$$

where N = number of available doctor hours per week

M = $N - 168$ = number of available doctor hours after allocating one doctor per hour of week

λ = $\sum_{i=1}^{168} \lambda_i$

The integer number of doctors to allocate per hour of week (r_t) is solved with a dynamic programming model which is then used to find the optimal number of doctors to allocate per shift, also solved by dynamic programming. The author notes that there may not be a feasible solution and sometimes there are multiple solutions. Therefore, a rough assessment tool was created by simulating an $M(t)/G/m(t)$ queueing system. Actual service times were noted as being difficult to measure and when simulated mean service time was more than 30 minutes, queues and wait times became "unrealistically large". The complexity of the service process was also noted including that service start time may not occur immediately following another patient's departure. This occurs since a new patient may be placed in a bed and then seen periodically by a physician. It is interesting that these challenges were noted in this early work as these are some of the persistent challenges receiving recent attention in current ED research.

In the United States (US), an early stochastic simulation model was developed by Saunders et al. (1989) at Vanderbilt University in Nashville, TN. At that time, the main purpose of the study was to show that complex features of EDs could be incorporated into a simulation model that could be run on inexpensive computer hardware and software. Their discrete-event simulation model was developed with the SIMAN language with supporting animation from the CINEMA package. Patient arrival times, triage acuity, performed tests

and procedures and diagnoses were input into the simulation model based on actual historical data from ED log sheets. The model adopts four levels of preemptive service, assigns each patient in the simulation to a specific nurse and physician, and incorporates multiple stages of service with different probability distributions applied to individual stations for tests, procedures and consultations. Output metrics from the model include patient wait times, queue lengths at various stages of service, staff utilization rates and patient throughput times. Simulated scenarios tested include varying the number of nurses, physicians, treatment rooms, and lab test times. Increasing the number of nurses or physicians showed a decrease in throughput time up to a certain point when no other decrease was shown. Increasing the number of treatment rooms did not decrease throughput times, presumably since the existing number was already adequate. Blood test turnaround times did show a direct effect on throughput time when the turnaround time was more than 60 minutes, but was insignificant when the turnaround time was below the 60 minute threshold. Suggestions for future applications include testing the effect of adding a fast-track for less emergent ED patients or simulating a community disaster when a sudden rush of serious and complex patient arrivals may occur.

In another early U.S. study, Tierney et al. (1986) investigated the extent that physicians can estimate the probability of myocardial infarction in ED patients with chest pain. During their physical examination, ED physicians completed a questionnaire that included the present complaint, past history, co-morbid diseases, medication history, and also their estimated probability of acute myocardial infarction. Logistic regression was used and the receiver operator characteristics (ROC) curve was plotted, which shows the ration of true-positive and false-positive rates. Physician estimates showed good prediction of myocardial infarction, and the area under the ROC curve is 0.87. While it was previously presumed that physicians would err on the side of caution (sensitivity or true positive rate) and admit

patients with low probability of myocardial infarction, the study results showed that physicians instead maximized the accuracy of patient classification (specificity or true negative rate).

Somoza and Somoza (1993) developed an artificial neural network to predict the admission decision in a psychiatric ED. Neural networks generally perform well in tasks that require pattern recognition and the judgment involved in the decision to admit or not is based on recognizing behavioral patterns in psychiatric patients. The one year study involves 658 of the 850 walk-in patients from the Department of Veterans Affairs medical center in Cincinnati, OH. The neural network is trained on the decision making process based on data collected from patient interviews. In order to determine if patient behaviors include patterns such as disorganized thinking or suicidal tendencies, data collected include features such as home, stressor, suicidal, brief psychiatric rating scale (BPRS), and primary emergency room diagnosis category (PEC).

Home refers to the number of people living with the patient, stressor is a scale of 1 (none) to 6 (catastrophic) of the stress the patient felt from their primary stressor, suicidal is measured on a scale of 0 (no thoughts of suicide) to 6 (suicide plan, e.g. left a suicide note), BPRS score is based on 18 items and ranges from 0 to 108 (converted into 6 levels). PEC has eight categories for the primary diagnosis from the ED. Many of these features are incorporated in the input later of the neural network.

The neural network is made up of three types of layers: the input layer, one or more hidden layers, and an output layer. In this case, the input layer includes features and the output layer represents the admission decision. The middle layer represents the interconnections between nodes and the strength of each connection is weighted. Weights are simulated initially with random numbers and an algorithm is used to train the neural network by gradually altering the weights in order to obtain a solution that minimizes the error between the clinician's decision and the neural network's decision.

The resulting accuracy of the neural network is fairly consistent with clinician decisions, and generally acts in a conservative fashion. Of the 271 patients that clinicians did not admit, the neural network's decision was not to admit 257 (94.8%) of those patients. As a result, the authors suggest that the neural network might be used as a screening mechanism and a psychiatric consultation could be requested to confirm if a neural network decision to admit a patient is valid or not. However, if the neural network were used instead of clinicians, 14 patients (4.4%) of the entire group would be discharged erroneously, and the authors question whether or not this is an acceptable level.

While it is well-known that the ED service process is complex, there are very few empirical studies examining the service process in more detail. Graff et al. (1993) conducted a time study of ED physician workload to test the hypothesis that physician service time varies by service category, LOS, and intensity of service. The study was conducted in a university-affiliated community teaching hospital with an annual census of 45,000 patients. Of the 12 physicians on staff in the ED, six participated in the study which measured service times for 1347 patients who received nonselected (514), walk-in (637), observation (52), laceration repair (102), or critical care (42) service respectively. Physicians recorded the beginning and end time for each service "interaction" with a given patient, and the total service time was measured by the sum of all interactions. Intensity of service was calculated as total service time divided by LOS. Prior to the study, service times were assumed to be similar for all patients regardless of service category and the American College of Emergency Physicians (ACEP) reported an average service time of 22 minutes per patient. The time study demonstrated that case mix does affect service time, and while service time did not vary significantly for nonselected patients (24.2 minutes) or laceration repair patients (25.0 minutes), it was significantly different for walk-in patients (9.8 minutes), observation patients (55.6 minutes), and critical care patients (31.9 minutes). The study also reported that walk-in patients and laceration repair patients typically had a single physician-patient interaction,

while observation patients had an average of 6.3 interactions and critical care patients had an average of 2.6 interactions.

Hoot et al. (2008) developed a discrete event simulation model at Vanderbilt University. The model aims to forecast ED crowding using a patient flow simulation approach that reports outcome measures including waiting count, waiting time, occupancy level, LOS, boarding count, boarding time, and ambulance diversion, all forecasted 2, 4, 6, and 8 hours into the future. Developed using the standard C programming language, the "ForecastED" simulation model incorporates:

A Time-varying Poisson arrivals for each hour of the day

B LWBS influenced by waiting room count at time of arrival

C Triage acuity level based on multinomial distribution

D Service times higher for "sicker" patients, log-normal distribution

E "Sicker" patients more likely to be admitted

F Poisson process for inpatient boarding

The model incorporates some of the complexity of the ED process, but excludes other important features including fast-track and a multi-stage service process with time for observation, consultations and lab tests.

2.7.1 Lab Tests

Point-of-care (POC) testing involves having laboratory tests completed in the ED rather than in a central laboratory. The goal is to reduce the turnaround time for laboratory test results providing the opportunity to decrease treatment times, the number of LWBS patients, and average overall ED LOS. POC testing can help reduce test turnaround times from 90 minutes to less than 10 minutes (Murray et al. 1999) and has become a practical option for EDs due to the miniaturization of biomedical devices. In a POC testing project at an urban academic hospital in Boston (Lee-Lewandrowski et al. 2009), the implementation of a blood quantitative D-dimer test in the ED reduced the test turnaround time from approximately 2 hours to 25 minutes. As a result, ED LOS declined from 8.46 to 7.14 hours and hospital

admissions were reduced by 13.8%. It is important to note that laboratory tests conducted in the ED may be more expensive than centralized laboratory tests. Therefore, in addition to clinical testing to ensure that test results do not reduce the quality of care provided, careful cost-benefit analysis should also be conducted. In cases where ED efficiency can be improved significantly, the benefits of POC testing should outweigh the added costs from having tests carried out in the ED rather than in a central laboratory.

Researchers from Vanderbilt University in Nashville, TN used a system dynamics simulation model to validate the effect of decreasing lab turnaround times on average LOS, daily throughput, and ambulance diversion (Storrow et al. 2008). The system dynamics modeled include stocks such as the number of patients waiting at triage, the number in the waiting room or the total number of patients in the ED, while differential equations are used to represent the flows of patients through the model. Patient data is collected from electronic patient records and tracking boards and input into a commercial simulation software (Patient Flow Center, Apogee Informatics Corp). In the model, ambulance diversion occurs whenever more than 10 patients are in the waiting room after more than 30 minutes of 100% ED bed occupancy.

The study was conducted in a large tertiary care adult ED with annual census of 55,000. Running the simulation model with 90 days of data, scenarios were completed to test the impact of decreasing lab test turnaround times of 120, 100, 80, 60, 40, 20, and 10 minutes respectively. Results demonstrated that as turnaround times decreased, ambulance diversion and ED LOS decreases while daily patient throughput increased. Note that these results are consistent with Saunders et al. (1989): LOS improves only if above a threshold (e.g. 60 minutes).

2.7.2 Generic Models

While many researchers work directly with one specific hospital to analyze ED patient flow, others explore the possibility of generic models for multiple settings. Sinreich and Marmor (2005) conducted a study of five hospital EDs in Israel to explore if there is a

unified process among EDs that can be integrated into a general simulation model. They suggest that such a model should be 1) generic and flexible, 2) intuitive and simple to use, and 3) initialized with reasonable default values for many of the system parameters. After examining the five EDs in the study, eight different patient types were identified: fast-track, internal, surgical, orthopedic, trauma, walk-in surgical, walk-in orthopedic, and internal/surgical. For each of the five EDs, patient flow process diagrams were developed for each patient type. Similarity measures were calculated and the process diagrams were found to be very similar. Therefore, a single unified patient flow process diagram was created incorporating several different elements and transitions including triage, initial examination, labs, imaging, consultation, treatment, and disposition to admit to the hospital or discharge home.

The authors also analyzed patient arrival data, found similar trends among hospitals in the study, and noted time-varying arrivals by hour of day and day of week. They conclude that patient type has a higher impact in defining the patient flow process than the specific hospital where the patient is treated. As a result, it is appropriate to develop a general simulation model based on the unified process.

Fletcher and Worthington (2009) completed a study to compare 'generic' and 'specific' emergency patient flow models for A&E, Bed Management, Surgery, Intensive Care, and Diagnostics. Models are classified into four levels: *generic principle*, *generic framework*, *generic model*, and *specific model*. An example of a *generic principle* is a theoretical queueing model that is not industry specific. A healthcare modeling toolkit in a simulation package is an example of a *generic framework* that is industry specific. Examples of *generic* and *specific models* are a simulation model for all A&E departments in the UK, or a simulation model for one specific A&E department. A more detailed set of dimensions is also presented including, for example, splitting generic models into models designed for central use or multiple local use.

The analysis is based on a literature review and e-mail survey to healthcare researchers, mostly members of the European Working Group on Operational Research Applied to Health Services (ORAHS). The literature review identified more evidence of specific models of A&E than generic ones. The authors note that model implementation is surprisingly rare, regardless of whether the model is generic or specific. In addition, studies that model connectivity between multiple departments or hospital wide models are much less common than specific department studies. In the case of A&E, the interaction with labs/imaging and (inpatient) bed management is usually limited to the impact that the processes have on A&E, rather than incorporating more detailed sub-models of those processes.

2.7.3 England's 4-Hour Rule

In England, a national target for A&E performance has been established by the Department of Health (DH). DH is responsible for England's publicly funded healthcare system, the National Health Service (NHS). The national target states that 98% of all A&E patients should be discharged, transferred or admitted within 4 hours of arrival. In December 2002, the national A&E average was 78% completed within 4 hours of arrival, so the 98% target would be a big challenge for many hospitals. In an effort to better understand specific challenges, DH developed a simulation model to help identify significant barriers to achieving the national target (Fletcher et al. 2007).

A generic simulation model was developed in Simul8 of a 'typical' A&E department. Although the Manchester triage system may have been adopted previously, the authors state that most hospitals had moved to simpler minor/major patient segmentation. As a result, only three generic patient flows were analyzed: minor, major, and admitted with the most complex flow for admitted patients.

Upon arrival, patients queue for an initial assessment by a doctor, followed possibly by diagnostic X-ray and/or blood tests and a second assessment. Treatment is provided followed by the admission disposition decision by a specialist. After the admit decision, patients remain boarding in hallways (referred to as "trolley wait" in England) until they

can be physically moved to the appropriate inpatient unit in the hospital. Note that while patients queue for assessments, diagnostic testing and admission are modeled as capacity unconstrained time distributions.

The simulation model uses time-varying arrivals by hour of day with inter-arrival times drawn from negative exponential distributions with process times modeled as simple triangular distributions. A variety of scenarios have been investigated with the simulation model and process times required to achieve the 98% target were identified.

In another study examining the feasibility of England's 98% target, Mayhew and Smith (2008) use a multi-stage queueing model to evaluate A&E completion times. In the initial model, patients are assumed to arrive according to a Poisson process at rate λ with exponentially distributed services times at each stage with parameter μ. Since average time spent at each stage is assumed to be the same, the probability of the total time in system equaling z is the sum of s random variables, and the probability density function of z follows a gamma distribution. However, such a model is not accurate for A&E, since the complexity of the service process is not adequately represented with the assumption that average service time is the same at each stage. In addition, service times are significantly different depending on, for example, if a patient is discharged home or admitted to the hospital. Therefore, the authors developed a more detailed model comprising three different treatment paths: "no/little", "short", and "long".

The no/little treatment path is modeled with a single stage exponential distribution, the short treatment path is modeled with a three-stage hypo-exponential distribution, and the long treatment path is modeled as a two-stage hypo-exponential distribution (later simplified to one-stage in a re-designated model). The authors use the re-designated model to examine the average completion times required to achieve various targets. To meet a 90% target within 4 hours, the required average completion time is 1 hour and 45 minutes. However, to meet a 98% target within 4 hours, the required average completion time is less than one hour. The authors argue that to achieve an average reduction of 45 minutes (43%)

represents a massive challenge for A&E. However, the model does not account for LWBS or time-varying arrivals. If peak and other variability in service requirements are considered, then improvements may be possible with flexible staffing policies.

After the implementation of England's mandate, shorter completion times have been reported. For example, in a sample of fifteen hospital trusts, the proportion of patients that were discharged, transferred or admitted within 4 hours increased from 83.9% to 96.3% from 2003 to 2006 (Mason et al. 2012). However, there has been a spike in the proportion of patients who completed within the last 20 minutes of the 4-hour target, particularly among elderly and admitted patients. The author's suggest that England's EDs "hit the target but missed the point".

2.7.4 Staffing

ED physician scheduling problems have been studied by applying both deterministic and stochastic operations research modeling approaches. Beaulieu et al. (2000) developed a deterministic mathematical programming approach to model the ED scheduling problem for the Sacré-Coeur Hospital in Montreal, Canada. Given a fixed number of physicians, planning period, and set of shifts, a multiple-objective integer program is used to identify an optimal ED physician schedule. Model constraints are classified as either compulsory or flexible, and any flexible constraints that are violated come at the cost of some "quality". The authors propose to order the constraints according to their relative importance, and formulate the model as a single-objective optimization problem that seeks to minimize the weighted sum of all deviations. Constraints of the model are classified in four categories: compulsory constraints, ergonomic constraints, distribution constraints, and goal constraints. *Compulsory constraints* include ensuring that all shifts are filled, physicians cannot be assigned to more than one shift per day, physicians assigned to a night shift cannot be assigned to a shift the next day, as well as vacations and particular shifts requested by physicians. *Ergonomic constraints* aim to set a more desirable overall schedule to individual physicians, for example, by limiting the number of successive working days. *Distribution constraints* take seniority

and other rules into consideration so that, for example, senior physicians work fewer weekend shifts. Goal constraints include setting a target number of hours per week to accommodate a specific physician's preference and fairly distributing night shifts among physicians.

Solving the model over a six-month horizon for 20 ED physicians in the study is not possible due to its dimension (40,000 variables with 75,000 constraints). Instead, six models of consecutive four-week periods are solved by branch-and-bound. However, due to conflicting constraints, no feasible solution exists so a heuristic iterative approach is adopted, incorporating branch-and-bound as a subroutine to iteratively improve the solution by satisfying more constraints.

In another study also conducted in Montreal, Carter and Lapierre (2001) interviewed ED physicians from six Montreal hospitals to try to gain a better understanding of the ED physician scheduling problem. Existing scheduling approaches are classified into three categories: *acyclic*, *cyclic without rotation*, and *cyclic with rotation*. *Acyclic* schedules are created separately for each period and were the most common type among the hospitals in the study. *Cyclic schedules* without rotation are schedules where physicians have the same shift patterns which repeat continuously. A *cyclic schedule with rotation* involves establishing a fixed number of schedules and each physician follow the same pattern (shift A, then shift B, and so on).

Similar to Beaulieu et al. (2000), Carter and Lapierre (2001) propose a model with "hard" and "soft" constraints. Of the six hospitals in the study, the authors describe their application of deterministic mathematical programming in two of the hospitals. In the case of Charles-Lemoyne Hospital, a revised cycle schedule is generated by solving the model by Tabu search. In the case of the Jewish General Hospital (JGH), acyclic schedules were necessary due to seniority rules and religious constraints. A revised schedule was proposed to the scheduler at JGH, with further refinements to be added for future "fine-tuning".

Another example of deterministic mathematical programming for ED physician scheduling is a study completed by Ferrand et al. (2011) for the Cincinnati Children's Hospital Medical Center. Cyclic schedules are created based on an integer programming model that takes into account factors such as holidays and vacation requests of individual ED physicians. The model is similar to Carter and Lapierre's general formulation for cyclic schedules, but in the Cincinnati case, each ED physician works a different number of hours, so individual schedules are developed to accommodate this requirement. The model incorporates constraints that address *regulatory constraints*, *work requirements*, and *physician preferences*. While *work requirements* are all hard constraints, some *regulatory constraints* and *physician preferences* are hard constraints while others are soft constraints. For example, the physician preference to never be assigned to more than two consecutive weekends is modeled as a hard constraint, while assigning no more than two consecutive overnight shifts during the Monday-Thursday period is modeled as a soft constraint.

The model was coded in AMPL and solved with CPLEX. An optimal solution for a set of five physicians was solved in less than six hours. In order to implement the model in the hospital, the authors also developed Visual Basic for Applications (VBA) macros in Excel to show the proposed schedule to the physicians. Three months after the implementation of the new cyclic schedule, physician feedback indicated that the new method provides "well-balanced" schedules and "relieves stress".

The ED staff scheduling models reviewed so far all deal with staffing only ED physicians, independent of other ED staff schedules. Other models examine the ED nurse staffing problem separately, such as Grano et al. (2008). However, altering schedules for physicians may impact the demand for nurses or other ED resources. Sinreich and Jabali (2007) introduce an iterative heuristic algorithm that combines both a simulation and optimization model to consider staff scheduling for multiple types of resources including physicians, nurses,

and imaging technicians. The methodology incorporates a staggered scheduling optimization model (S-model) and an iterative simulation based Staggered Work Shift Scheduling Algorithm (SWSSA).

The S-model objective is to minimize the sum of all overstaffing and understaffing penalties. The authors suggest that while understaffing is generally not appropriate in the ED, an assumption is made that nurses and physicians can be shared between the ED and hospital wards, whereby the understaffing penalty cost is incurred. The understaffing penalty is suggested to be greater than the overstaffing penalty as it is also assumed that transferring available resources to a hospital ward is easier than receiving resources from a hospital ward to work in the ED.

In each iteration of the SWSSA, eight weeks of data are generated by running three replications of the simulation model by Sinreich and Marmor (2005). In the initial iteration, the model is solved with the initial schedule of all resources. A weighted average LOS is calculated along with a patient flow delay factor, for each resource type. The resource with the largest delay factor is identified as the critical resource (bottleneck) at each stage of the algorithm, and the capacity of the critical resource is increased to a large value for the next stage of the algorithm. Improvements are iteratively obtained for each resource type as the critical resource changes from one iteration of the algorithm to the next.

The algorithm was tested using data from five Israeli hospitals. Results show that physician and nurse hours can be reduced by 8-17.5% and 13-47% respectively, while maintaining LOS values within -19 to 4% of the original values. The authors conclude that selective downsizing of an ED workforce is possible without impacting ED efficiency. However, the model does not account for overtime and other flexible staffing costs. Furthermore, the assumption that all physician and nurse resources can be shared between ED and inpatient units is not realistic. In practice, specialty ward inpatient care cannot be handled by ED physicians since specialists are the only physicians with the appropriate training to care for those patients.

Using similar methodology that also combines simulation and optimization models, Sinreich et al. (2012) introduce iterative heuristic algorithms with the objective to reduce patient wait times by leveling resource utilization. Considering that ED patient care is provided by physicians, nurses, imaging technicians and other health care professionals over the course of several hours, arrival time is not a good indicator of when all heath care personnel are needed. The authors develop a work shift scheduling mixed-integer programming optimization model called *Sched-Opt*. The Sched-Opt objective attempts to level the ratio of available to required resources units throughout the hours of the day. The purpose of Sched-Opt is to identify the optimal starting times for a weekday. One of two iterative heuristic work shift scheduling algorithms, WSSA-1 or WSSA-2, is then used to identify optimal work shifts for each resource type. The algorithms are similar, except time blocks can be transferred from one nurse type to another and one physician type to another in WSSA-2, but not in WSSA-1. Using delay factors for each resource for each hour of the day, a critical resource is identified, and the simulation model from Sinreich and Marmor (2005) is invoked at each stage of the algorithm. Work shift schedules are improved for the most critical resource at each stage. Improvements are essentially achieved by shifting resource capacity from non-peak periods to peak periods, resulting in more efficient service for ED patients who are treated during peak periods.

Tests for WSSA-1 and WSSA-2 were completed with data from five hospitals in Israel. Greater improvements were observed in medium and large hospitals. WSSA-1 resulted in reduced wait times of 20-45% and reduced LOS of 7-17%, while WSSA-2 resulted in reduced wait times of 20-64% and reduced LOS of 11-29%.

In another example of a simulation optimization approach for ED staffing, Ahmed and Alkhamis (2009) determine the optimal number of doctors, nurses, and lab technicians for a government hospital in Kuwait. Considering budget restrictions and resource constraints, the authors investigate staffing allocation that maximizes patient throughput and reduces patient wait times.

The simulation incorporates time-varying arrivals according to a non-homogeneous Poisson process for three patient categories according to acuity level (category 1 patients are critical and categories 2 and 3 are non-critical). After reception, patients wait for an available examination room. Patient acuity is determined in the examination room by a doctor (rather than by a triage nurse), and lab tests are requested if necessary. After assessment, non-critical patients either wait for minor treatment provided by a nurse (category 2) or receive medication and are discharged (category 3). Critical patients (category 1) stay under close observation and are treated by a nurse and a doctor who is called in to the examination room whenever needed. Some critical patients are discharged and others are admitted to an inpatient unit in the hospital for further care. The admission rate of all patients in this ED is 12%. Service time distributions for lab tests are modeled with a triangular distribution, while other stages including reception, doctor examination and reexamination, nurse treatment times are modeled with uniform distributions.

In the optimization models, resource constraints are included for each resource type x_i, based on restrictions on physical layout and other factors identified by ED administrators. In this setting, staffing levels cannot exceed three receptionists (x_1), four doctors (x_2), five lab technicians (x_3), six treatment room nurses (x_4), and 12 emergency room nurses (x_5). Two separate optimization problems are formulated, one with the objective of maximizing throughput (A-1) and another with the objective of minimizing cost (B-1). Problem A-1 is a discrete stochastic optimization problem with one stochastic and two deterministic constraints. A two phase procedure is used to solve the constrained stochastic optimization problem. In the first phase, a set of solutions satisfying the deterministic constraints is first determined. Next, this set is updated by setting lower and upper bounds on Q1 in order to eliminate solutions that violate the stochastic constraint. In the second phase, the best among the remaining solutions is selected by an optimization algorithm with stopping criteria for a specified number of iterations or when the solution has not improved after a specified number of iterations.

Problem B-1 is a deterministic optimization problem with two stochastic constraints, and one deterministic constraint. A two phase procedure is also used to solve problem B-1. However, this time the feasible solution set is found using a different feasibility detection procedure that handles the case of more than one stochastic constraint. Once the set of possible solutions is determined, identifying the best solution among them is simply the solution to a deterministic integer programming problem that can be solved easily with any solver such as the MS Excel solver add-in.

Most of the studies reviewed so far focus on models to assist ED management in testing interventions *offline* before implementation. Models that support *real-time* ED operations management have also been developed by Zeltyn et al. (2011) in Israel. The focus of the paper is on simulation models to address ED staffing challenges over operational, tactical, and strategic horizons. At the core of the article is an offered load approach for short-term operational staffing decisions. Tactical problems involve accommodating seasonal factors such as increased arrivals due to the flu. Strategic problems involve planning for major design changes such as a physical relocation of the ED within the hospital. We focus our review on the proposed methodology for operational staffing since it is the core of the paper.

The study is based on field research in nine Israeli EDs with a specific focus on the large government-affiliated Rambam Hospital, a medical center that serves over 2 million citizens (one-third of Israel's population) with approximately 82,000 ED patients per year. ED patient types are classified in one of six categories: 1) internal acute, 2) internal walking, 3) surgical acute, 4) surgical walking, 5) orthopedic acute, or 6) orthopedic walking. The walking patients can use chairs, while acute patients require an ED bed. The methodology used for operational staffing has the following steps:

1. Obtain the initial, current ED state by simulation
2. Generate stochastic patient arrivals
3. Run the simulation model with *infinite* staffing resources for eight simulated hours
4. Calculate staffing recommendations using an offered load method

5. Run the simulation model from the current ED state with the recommended staff
6. Calculate performance measures

In step 1, simulated data is necessary since the current ED state is only partially captured from hospital information systems. Improvements to data accuracy and completeness may come in the future if real-time tracking systems are successfully adopted. In the mean time, actual arrival data is fed into the simulation system which is then used to generate simulated arrivals that are consistent with real arrival patterns. An appropriate initial state is generated after a simulation warm-up period of three weeks. In step 2, forecasted ED arrivals are based on long-term moving averages (MA). Considering that arrival rates vary by both time of day and day of week, long-term MA are calculated based on historical data for each hour of the week to estimate the arrival rate.

The rest of the methodology relates to the service process and the offered load approach. While other methods, such as rough cut capacity planning (RCCP), ignore the time lag between arrivals and the time when service is required, the offered load approach spreads workload more over time. In EDs, this time lag is significant, and as a result, arrival rates will reach maximum often before resource workload reaches maximum. In the simplest case of an M/M/1 queueing system with arrival rate λ and service rate μ, the offered load is $R = \frac{\lambda}{\mu}$. Staffing rules can then be determined in terms of R, such as the square root staffing rule (Jennings et al., 1996):

$$n = R + \beta\sqrt{R}$$

where β is set to ensure a desired service level. The square root staffing rule is commonly used within the quality and efficiency-driven (QED) regime, for service systems that require both high service quality and high resource utilization. Considering that the ED has time-varying demand, a modified offered load (MOL) approximation is used and the square root staffing function is replaced with a time-varying square root staffing function (Feldman et al., 2008):

$$n(OL,t) = R(t) + \beta_t \sqrt{R(t)}$$

The offered load approach involves first running the simulation model with infinite staffing capacity (step 3). For each resource type, the number of busy resources required during each hour is determined from the infinite staffing simulation run, which is used to estimate $R(t)$ for each resource type. The MOL approximation of $R(t)$ is obtained from averaging the results of multiple simulation runs, and the time-varying square root staffing function is calculated to determine the recommended staffing levels for each resource type during each hour (step 4). The recommended staffing levels are then used to run the simulation again (step 5) and performance measures are calculated (step 6).

Izady and Worthington (2012) propose a similar methodology (that also uses a MOL approach with infinite server networks, square root staffing, and simulation) to determine the minimum staffing levels required to meet England's 4-hour target. Three types of patients and six staff resource types are considered. Patient types are minor, major, or admitted; resource types are doctors, emergency nurse practitioners (ENPs), electrocardiogram (ECG) technicians, lab technicians, radiologists, and nurses.

The patient flow process for minor patients is assumed to begin with a first assessment by a doctor or ENP, and treatment is assumed to be completed by a nurse. In some cases, treatment is preceded by diagnostic testing (ECG, lab, or radiology) which is followed by a second assessment before treatment. All minor patients are assumed to be discharged home after treatment. Major and admitted patient types are assumed to follow a similar process, except assessments are only conducted by a doctor and a resuscitation room is available for patients who arrive with serious conditions. Multiple doctors may be allocated for patients who require a resuscitation room. All major patients are assumed to be discharged home and admitted patients are all assumed to be admitted to a hospital ward for further care.

The authors consider an $(M(t)/G/s_k(t))^K$ queueing network with service stations $k = 1, 2, , K$. A stationary Markovian routing process is assumed throughout the network, with

fixed probabilities assigned to different routes for the minor, major, and admitted patient types. Using a MOL approach, an estimate of the offered load, $m_\infty^k(t)$, is obtained from the mean number of busy servers for each service station (resource type) k from solving the corresponding infinite server $(M(t)/G/\infty)^K$ network. For each resource type, the time-varying square root staffing function is then applied to determine the recommended number of servers, $s_k(t)$:

$$s_k(t) = m_\infty^k(t) + \beta\sqrt{m_\infty^k(t)}$$

where β is a quality of service parameter chosen according to the targeted delay probability α. Note that based on the heavy traffic limit theorem (Halfin and Whitt, 1981), the relation between α and β is :

$$\alpha = \left[1 + \beta\frac{f_1(\beta)}{f_2(\beta)}\right]^{-1}$$

where f_1 and f_2 are the density and cumulative distribution function (cdf) of the standard normal distribution.

The staffing algorithm for achieving the 4-hour target for 98% of patients is as follows:

1. Set $\alpha = 1$ and calculate $m_\infty^k(t)$ for each resource type k during each staffing interval t.
2. Find β.
3. Calculate $s_k(t)$ using the square root staffing function for each resource type k during each staffing interval t.
4. Given staffing levels from step 3, run a simulation to estimate the percentage of patients with completion times within four hours
5. If percentage is less than 98%, decrease α and go back to step 2

The authors tested the approach on a 'typical' A&E department in the UK from Fletcher et al. (2007). The baseline staffing profiles, based on average resource utilization from the Fletcher et al. (2007) simulation model, achieved completing times within 4-hours for 96% of patients. After six iterations of the algorithm, the 98% target could be achieved by

staffing according to a delay probability $\alpha = 0.75$, corresponding to $\beta = 0.221$. Comparing the resulting balanced staffing profiles to the baseline staffing profiles showed more stable resource utilization over time. As a result, the 4-hour target can be achieved with the same number of doctors, one hour more of ECG technicians, and fewer hours of ENPs, lab technicians, radiologists, and nurses.

However, after determining the desired staffing profiles, the authors note that feasible shift schedules may not coincide with the balanced staffing profiles. Therefore, an integer programming approach based on Sinreich and Jabali (2007) is used to determine shift schedules that are as close to the balanced profiles as possible. The results show that shift scheduling constraints reduce the amount of staff savings, but the 4-hour target can still be achieved, while using fewer staff hours than the baseline staffing profiles.

2.8 Admission & Boarding

Patients are admitted to hospital wards through direct admissions from clinics for scheduled procedures, transfers from other healthcare facilities, or the ED. From an ED patient flow perspective, the main bottleneck is the inpatient exit rate which is generally outside the control of ED management. While patients are waiting in the ED for an inpatient bed, they are often moved out of the ED bed but remain boarding in a hallway which limits some ED resources for new patients. During periods of peak congestion, some patients elect to leave without being seen (LWBS) by a physician which can have negative consequences on both patient safety and satisfaction. Patient dissatisfaction and ambulance diversion from overcrowded EDs creates a loss of demand for ED services (Schull et al. 2001). Oddly, this situation actually may be financially beneficial to healthcare organizations. Considering that most of the expenses incurred by hospitals are fixed costs, hospitals seek occupancy rates of 100%. The current system allows hospitals to maximize their profits by operating at full capacity while keeping an overflow of in-patient demand waiting in the ED to be admitted. While lost ED revenue seems undesirable, losing or deferring admitted patient revenue from other sources may be less desirable (Handel et al. 2010). However, Pines et al. (2011) argue

that dynamic bed management strategies can be developed to reduce inpatient boarding in a way that is financially beneficial for hospitals.

ED LOS is particularly long for admitted patients who are stuck boarding in hallways (for several hours or even days), which is clearly a concern from a patient satisfaction perspective. Pines et al. (2008) reviewed patient satisfaction surveys and showed that prolonged boarding times are associated with low satisfaction. Furthermore, it has been shown that admission delays also increase inpatient LOS and inpatient cost for the admitting unit. In a study of two Canadian hospitals, admitted patients whose ED LOS exceeded 12 hours resulted in 12% higher inpatient LOS and 11% greater inpatient cost (**?**).

Jones et al. (2002) developed a forecasting model to predict the number of inpatient beds occupied due to emergency admissions. Using six years of data from a UK hospital, the authors observed both monthly and weekly seasonality. Maximum values were observed during winter months, while minimum values were observed during summer months. Bed occupancy levels due to ED patients were highest on Mondays, declining until Thursday with a small increase on Friday, followed by a decline over the weekend. Using Seasonal Auto-Regressive Inductive Moving Average (SARIMA) modeling, a relationship was found between the number of occupied beds and two variables: mean daytime temperature and the influenza illness rate.

Good forecasts are produced with SARIMA most of the time, but the model fails to produce good forecasts during a bed crisis. The authors also explored the use of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) modelling to account for the volatility from cancelled surgeries and ED congestion. GARCH is used for time series forecasting when periods of high volatility, followed by periods that are relatively stable. The authors found that periods of high volatility can result in pressure for inpatient beds for up to fourteen days. However, increases in ED waiting times due to inpatient bed occupancy resulted in much shorter lags of five days or less. Therefore, the authors state that ED waiting times for inpatient beds are attributed more to volatility than to bed occupancy.

To gain a better understanding of the influence of volatility on ED wait times, the authors suggest further research using other methodologies such as system dynamics and discrete event simulation.

Lane et al. (2000) developed a systems dynamics simulation model to investigate how changing the number of inpatient beds affects A&E waiting times. The study was conducted in a U.K. teaching hospital (referred to as "St. Dane's" for confidentiality reasons) through collaboration between the University of London and the NHS. System dynamics is used to examine the interaction between A&E and inpatient wards by considering aggregate system-level stock and flow variables. The model was built using iThink software on a Macintosh computer. The model considers two main patient groups: emergency admissions and elective patients. The arrival rate of emergency patients to A&E and the rate of scheduled elective admissions were modeled as exogenous variables, based on historical data. Elective patients include both surgical patients for procedures such as hip replacement, as well as medical patients for services such as chemotherapy. Emergency patients are typically given priority over elective patients, so pre-scheduled elective admissions are often cancelled as inpatient bed occupancy rises.

The model's causal loop diagram includes 1) loops for patients that depart A&E: discharge or admit to hospital ward, 2) an inpatient bed occupancy level loop: restricting emergency admissions or elective admissions when all beds are full and 3) loops to reduce the backlog of scheduled elective patients, either by admission or by cancellation. The core of the model has nine stock and 160 flow variables. For example, the number of Scheduled Elective Admissions is a stock variable affected by flow variables for the initial Scheduling Rate, along with the Drop Out Rate and Elective Cancellation Rate.

Simulated scenarios include changing A&E demand and inpatient bed capacity. Considering that the government had hypothesized that reducing the number of hospital inpatient beds would not affect the level of service provided to emergency patients, reducing inpatient bed capacity was an important scenario to test. The results from this scenario did

counter-intuitively support the government's hypothesis, as simulated waiting time in A&E did not vary significantly by varying bed capacity between 700 and 900 beds. However, while reducing the number of inpatient beds may not increase A&E waiting times, the number of cancellations for elective patients is increased significantly. In the scenario with 100 fewer beds, in order to achieve the same A&E wait time, the number of elective cancellation doubles compared to the base case. Note that this situation is apparent in the model's causal loop diagram, corresponding to shifting flow from elective admissions to elective cancellations.

Scenarios to test changes in A&E demand include permanent changes in demand as well as surges from a crisis event. Small changes in permanent demand resulted in reduction in the number of elective cancellations, but to a lesser extent than when inpatient bed capacity is reduced. However, small changes in permanent demand resulted in a large impact on A&E patients: a 4% demand increase resulted in a 45-minute increase in average LOS. The crisis event simulation involved a 13% demand surge, modeled based on a real surge event experienced at St. Dane's previously. The results showed a 5 day period after the surge before the system returned to normal operating levels. The simulated results were consistent with the surge event that St. Dane's had observed previously. In the scenarios when the system breaks down due to A&E demand increases, the key bottleneck within A&E was the first physician assessment.

Other factors influencing A&E waiting times for admitted patients were also noted. A nurse and porter are required for patient transport between A&E and the inpatient ward. When A&E is crowded, delays may occur if all nurses or porters are busy. Bed turnover time may also vary depending on inpatient staff workload. These delays or other factors could result in elective patients being assigned inpatient beds even when an emergency patient may be waiting in A&E with a higher priority. Note also that since systems dynamics models consider aggregate flows, the stochastic variation in processes at the individual patient level is not captured.

Recent models have specifically investigated the effect of inpatient boarding on ED efficiency. Khare et al. (2009) used discrete event simulation to model ED patient flows for an urban, academic, tertiary care, Level I trauma center in Chicago that receives over 75,000 adult ED patient visits per year. The simulation model was built using MedModel (ProModel, Orem, UT) and was used to compare the effect of two different interventions on average ED length of stay (LOS). The study showed that increasing the number of ED beds did not reduce LOS but increasing the rate that admitted patients depart the ED did result in a significant reduction in LOS.

The model assumes Poisson patient arrivals with peak and non-peak periods. Patient flow varies after triage, based on ESI levels. ESI 1 patients skip the queue and go directly to the Main ED, while other patients sit in the waiting area and queue until an ED treatment bed is available. ESI 4 and 5 patients wait for the fast-track Urgent Care unit, while ESI 2 and 3 patients wait for the Main ED. Urgent Care patients are assumed to have an initial time with a physician, followed by a treatment time that does not require a physician. All urgent care patients are assumed to be discharged home from the ED. On the other hand, Main ED patients are assumed to have a more complicated service process, which begins with an initial time with a physician, followed by a treatment time, and a second physician visit before disposition: discharge or admission & boarding. For admitted patients, boarding time is modeled with an exponentially distributed inpatient exit rate.

For each ESI level, LWBS patients were modeled using a threshold time that a patient will wait before leaving without being seen and a probability of LWBS. The authors assumed that no ESI 1 or 2 patients left without being seen due to the severity of their condition, but ESI 3, 4, and 5 patients left without being seen based on two rules: 1) 25% of ESI 3 leave if not seen by a physician after 90 min and 2) 50% of ESI 4, 5 leave if not seen by a physician after 60 min. The base case simulates a 23-bed Main ED with an exponentially distributed admitted patient departure rate of one patient leaving the boarding area of the ED every 20 minutes. The authors tested scenarios including 1) increasing the number of

Main ED beds to 28 and 2) increasing the admitted patient departure rate to one patient every 15 minutes. The results showed that increasing the number of Main ED beds did not decrease average LOS. However, increasing the admitted patient departure rate did result in a significant reduction in average LOS from 240 to 218 minutes. The robustness of the results were tested by performing sensitivity analysis that varied a number of parameters including changes to daily census, patient mix, LWBS rates, treatment times, admission rates, and further changes to the number of ED beds and admitted patient departure rates. The authors found similar trends and concluded that the rate that admitted patients depart the ED is the main bottleneck in ED patient flow.

Bair et al. (2010) also developed a discrete event simulation to investigate the effect of inpatient boarding on ED efficiency. The study was based on a Level I academic trauma center in California with approximately 60,000 ED patient visits per year. The authors used NEDOCS and the rate of LWBS patients per day to assess the degree of ED efficiency in a setting with separate adult and pediatric units. The study investigated the effect of altering the boarder-released-ratio (r = ratio of admitted patients that do not wait to board compared to all admitted patients). The results showed a significant decrease in ED crowding and LWBS after altering the boarder-released-ratio from 0% to 100%.

Inter-arrival times are modeled with a shifted beta distribution with 12 random number streams, one every two hours throughout the day. The triage process in the model involves segmentation first between adult and pediatric patients as well as further classification as one of five triage levels (red, orange, yellow, green, or blue). Patients are prioritized based on triage level and queue for treatment. Pediatric patients are sent to one treatment area (Area 2), while adult patients are sent to one of three treatment areas (Area 1, 3, or fast-track). Treatment is non-preemptive and treatment times are fitted to a shifted beta distributed function. A proportion of patients are discharged from the ED while others are admitted. Boarders are processed first-in-first-out (FIFO) and inpatients are segmented into ICU and non-ICU inpatients.

LWBS patients were modeled in the simulation by applying a probability that an arriving patient leaves without being seen after triage. More patients will leave without being seen during peak periods when wait times are higher due to ED congestion. To incorporate the degree of ED crowding, the departure probabilities used in this study were calculated as a function of the NEDOCS score at time t. Results based on 10,000 simulated days showed that the boarder-released-ratio had a significant effect on LWBS and ED crowding. Scenarios involved increasing the boarder-released-ratio in 10% increments from 0 to 100%. For example, if 50% of boarders could be admitted directly to an inpatient ward without boarding, the proportion of overcrowded days according to the NEDOCS score would decrease from 88% to 68%.

Ceglowski et al. (2006) consider an ED modeling approach that combines data mining and discrete event simulation (DES). The data-driven model uses non-parametric methods to identify the groups of medical procedures provided to ED patients. Each group of procedures represents a distinct treatment pathway. The authors identified 20 different *treatment pathways*, and when considering a 5-level triage system, 4 different disposal codes, and patients in a separate "dead on arrival" category, there are 401 possible *patient types*. However, in practice many of the patient types occurred rarely or not at all, with 99% of patients belonging to one of 161 different patient types.

In the DES model (developed in Simul8), the state of ED treatment sites is tracked as either occupied are free. Queues form whenever all treatment sites are occupied. The model results show that the heaviest users of ED beds are patients waiting for admission to a hospital ward. For those patients, the decision to admit was made earlier in their treatment, but remain in the ED due to admission delays. The authors note that the treatment and symptoms of patient records give an indication of the corresponding wards with admission delays.

The authors conclude by suggesting the following general simple methodology for identifying ED to ward bottlenecks (without needing simulation): 1) Gather and prepare data,

2) Divide patients into homogeneous groups and calculate average bed times, 3) Conduct an analysis of bed time and the number of patients in each patient group, identifying a utility function of the demand placed on the system for each group. This simple analysis can identify the most critical ED to ward interactions. However, the authors note that experimenting with the utility function may produce inaccurate results without a more sophisticated model due to the non-linear complexity of the ED system.

In Ireland, Abo-Hamad and Arisha (2013) developed a discrete event simulation model of a 570-bed adult teaching hospital that serves over 55,000 patients annually. In 2007, the hospital's ED performance measures included an LWBS rate of 17%, and average LOS over 9 hours with standard deviation of over 3 hours, which is much longer than the national 6-hour target. The authors combine multi-criteria decision analysis (MCDA) tools along with simulation and a balanced score card (BSC) to illustrate that unblocking ED outflows by inpatient bed management is more effective than increasing the physical ED space or ED workforce.

Many ED studies focus on one performance measure such as LOS, however, in practice ED managers rely on a variety of key performance indicators (KPIs). KPIs for ED performance, as specified by senior hospital management, include patient throughput KPIs such as waiting time and LOS for admitted and discharged patients, as well as ED efficiency KPIs such as resource utilization (staff or capacity), layout efficiency (walking distances) and ED productivity (% treated patients).

After detailed processing mapping, the authors built a simulation model using object oriented programming with modules for the corresponding processes. Modules include patient arrival, registration, triage, patient allocation, patient treatment, and patient transfer. Sub-processes are also created for more detailed resource requirements such as staff and equipment required at each point of care. For example, patient treatment requires medical staff, medical equipment, bed/trolleys/seats and cubicles and includes the sub-processes: seen by a doctor, referred for opinion, referred for admission, and awaiting admission.

Simulated scenarios include: 1) a zero-tolerance policy for exceeding the national 6 hour target, 2) increasing the number of trolleys by 50%, and 3) adding one doctor to the overnight shift. Scenario 1 would involve either improving the inpatient admission and discharge cycle in the hospital or moving boarders to a short stay unit. Admitted patient LOS reduced from 21 hours to less than 8 hours. Scenario 2 created more physical space and decreased the number of patients in the waiting room, but increased the number of admitted patients and resulted in a 5% increase in average LOS. Scenario 3 had a significant reduction in queue length and wait room time reduced by 44%, and average LOS reduced, but not enough to meet the national 6-hour target. The authors infer from the results that investing in improving the admission/discharge cycle within the hospital is the most effective strategy. Additional scenarios based on combinations of scenarios 1-3 were also tested, and tradeoffs between competing PKIs were noted by the authors. The preferences of ED management for weighting the PKIs was unclear, so the authors used preference ratios in multi-attribute evaluation (PRIME), a methodology to handle MCDA when there is incomplete information about decision maker preferences.

In a Canadian study of the Toronto General Hospital, Wong et al. (2010) built a system dynamics (SD) simulation model to examine the effect of smoothing inpatient discharges from the general internal medicine (GIM) ward. The simulation model results demonstrated that ED congestion could be significantly reduced if hospital discharges of GIM inpatients could be spread evenly over the course of the week. In the study hospital, 98% of GIM patients are admitted from the ED and represent almost half of the total number of inpatient admissions from the ED. Bed blocking is also prevalent due to GIM patients with approximately 25% of GIM patients occupying beds while waiting for alternate level of care (ALC) facilities. The model was built using one year of historical data (2005) and validated with a second year of data (2006).

GIM patients are either discharged from the ED or transferred to an inpatient ward, with three main patient pathways considered for inpatients: Discharge Home, Discharge Inter-Facility, and Discharge Other. Discharge Other includes intra-facility transfers, in-hospital deaths and patients who leave against medical advice. Corresponding stock variables include the number of GIM patients in the ED, as well as the number of GIM home, inter-facility, and other patients respectively. Three endogenous functions are included: 1) Discharge from ED Probability, 2) Bed-Turn-Around Time and 3) ALC Occupancy. *Discharge from ED Probability* is modeled as a function of GIM patient occupancy in ED which occurs when boarding times are so long that patients stabilize and discharge from the ED. *Bed-Turn-Around Time* is modeled as a function of the number of ward discharges, including the time required for housekeeping staff to prepare bed and for nursing to accept a new patient. *ALC Occupancy* is approximated in the model based on stock of inter-facility patients. Note that ALC patients occupy an inpatient bed while waiting for space in long-term care, rehabilitation care, or other facility.

Tested scenarios include a smoothed average case and an "every day is a weekday" case. In the smoothed average case, demand is smoothed using the 7-day weekly average discharge proportion. The "every day is a weekday" is similar, except that the 5-day weekday average discharge proportion is used. In the smoothed average case, the GIM in ED stock variable reduced from 7.4 to 5.4 (27% decrease), and average ED LOS for GIM patients reduced from 24 hours to 17 hours (31% decrease). Larger reductions were shown with the simulation results for the "every day is a weekday" case: GIM in ED reduced by 48-57% and ED LOS for GIM patients reduced by 51-60%.

While the potential benefits of smoothing inpatient discharges over the course of the week are clear, there are important implementation challenges. Weekend inpatient discharges are less common that weekday inpatient discharges not only due to ward staff availability, but also due to resource capacity of external facilities and for patient safety considerations. Transfers to long term care and other facilities may simply not be possible since few of

those facilities accept new patients on weekends. Patient safety concerns may also reduce the ability for patients to be discharged home, since less community and other resources are available on weekends. Also note that the model uses average weekly rates for hourly admissions and discharges rather than day-of-week rates, and that the inpatient discharge pathway (home, inter-facility, or other) is assumed to be known when transferred out of the ED.

In addition to day of week inpatient discharge timing, time of day inpatient discharge timing also affects ED congestion. Powell et al. (2012) examined inpatient discharge timing and found that discharging more inpatients from the hospital earlier in the day could significantly reduce and possibly eliminate ED boarding altogether. In a retrospective analysis, pre-existing patient records were extracted from hospital information systems for the month of September 2007. The analysis included weekday inpatient bed demand from three sources: *ED Admission*, *Elective Surgical Admission*, or *ICU Transfer Admission*. The authors developed a simplified model of the interaction between the ED and inpatient units, based on the average hourly mean values of the three sources of inpatient demand (inflow) and inpatient discharges (outflow). Model assumptions include: priority for elective surgery patients over ED patients, priority for ED and elective surgery patients over patients transferred from the ICU, and no bed specialization (any unit can take care of any type of inpatient admission).

In the hospital's current practice, 70.5% of ED patients, 85.6% of elective surgical patients, and 82.8% of ICU transfer patients are admitted to the hospital between noon and midnight. Inpatient discharge timing policies include "Discharge by Noon", "Dayshift Uniform Discharge", and shifting the hospital's actual discharge distribution curve 1, 2, 3, or 4 hours earlier. Specifically, the "Discharge by Noon" policy is modeled such that 75% of inpatients are discharged uniformly between 8:00 AM and 12:00 noon, with the remaining 25% discharged uniformly between 12:00 noon and 8:00 PM. In the "Dayshift Uniform Discharge" policy, inpatients are discharged uniformly from 8:00 AM to 4:00 PM (corresponding to normal inpatient physician working hours).

The primary outcome measure is total daily admitted patient boarding hours. In the model base case, the total boarding hours per day is 77: 56.3 hours for ED patients and 20.7 hours for surgical patients. Shifting the hospital's actual discharge distribution curve by one hour earlier eliminated surgical patient boarding and reduced ED patient boarding to 34.4 hours per day. To eliminate ED boarding altogether, the curve would need to be shifted four hours earlier. The "Discharge by Noon" and "Dayshift Uniform Discharge" policies resulted in eliminating surgical patient boarding and reduced total ED patient boarding to 3 hours per day. The authors note that future work could identify physician and nurse staff shift schedules required to achieve the proposed or other optimal inpatient discharge policies under different hospital-specific constraints. Recently, Shi et al. (2015) developed a stochastic processing network model of the admission process. Their simulation studies also showed that improvements could be obtained through early discharge strategies. They suggest that a patient's medical needs determine the required number of nights for their hospital stay, but operational policies affect time of day admission and discharge timing.

In a study focused on ambulance diversion, Allon et al. (2013) propose a two-station queueing network model of the interface between the ED and inpatient ward. The first station represents the ED and the second station represents the inpatient department. In their model, patients arrive to the ED by walk-in or ambulance according to a Poisson process with priority given to ambulance patients. ED beds are modeled as servers and after receiving exponential-distributed service in the ED, a fraction of patients are admitted to the hospital while the rest are discharged home from the ED. For the inpatient department, patients arrive from the ED or directly to the inpatient department according to a Poisson process. Priority to ED patients is always assumed and inpatients beds are modeled as the servers with exponential-distributed service times. When all inpatient beds are full, it is assumed that ED boarders occupy ED beds until an inpatient bed is available. In practice, ED boarders often wait on stretchers in hallways for admission to the inpatient wards. Further assumptions are made and the model is reduced to two separate single-station queueing

systems. The simplified system is then analyzed using heavy traffic approximations and applied to hospitals in California. The study findings highlight the fact that ED crowding is a multi-departmental problem: the extent that a hospital goes on diversion varies on the size and occupancy of both the ED and inpatient departments.

Mandelbaum et al. (2012) modeled the interface between EDs and internal wards (IWs) in a large Israeli hospital as an inverted-V queueing system with a single centralized queue and multiple heterogeneous server pools: the pools represent the wards and servers are beds. The authors analyze patient allocation from the ED to IWs from the perspective of fairness towards ward staff (rather than fairness towards patients). The ED-to-IW process is analyzed within the quality-and efficiency-driven (QED) regime (Halfin and Whitt 1981) applicable for mid-to-large scale queueing systems that require both high levels of service quality and resource efficiency. QED is natural in this setting since wait times (ED boarding times, measured in hours), are significantly shorter than service times (average LOS in wards, measured in days) and high bed occupancy is the norm. Note that while the servers in the model are heterogeneous since the number of beds in each ward may be different, all of the wards have similar medical capabilities and offer similar services. This is common in Israeli hospitals but may not be applicable elsewhere.

The model considers a single customer class and assumes that ward arrivals occur according to a Poisson process with exponentially distributed ward LOS. As noted by the authors, these assumptions are inaccurate in practice, but are included in order to characterize a tractable analytical model. The queueing discipline is first come first serve (FCFS), nonpreemptive (service cannot be interrupted once started) and work conserving (no idle servers when patients are in the queue). Three routing algorithms are considered, all with the work conserving goal of reducing the length of the queue of patients boarding in the ED. As the name suggests, the longest-idle-server-first (LISF) policy routes the next customer to the server that has been idle for the longest. This policy is commonly used in call centers and considered fair (Armony and Ward 2010); however, implementation is not trivial

in hospitals. To implement LISF, hospitals would need to keep track of the number of idle beds in each ward, ordered by idle time. Queues-and-idleness-ratio (QIR) policies (Gurvich and Whitt 2009) were also considered where the next patient is routed to the server pool with the highest number of idle servers. While this policy can be used to achieve the same level of fairness as LISF, it is also not straightforward for the hospital setting, due to pool capacities varying over time. Therefore, a new policy was introduced in the paper, named the randomized most-idle (RMI) routing policy. In RMI, a patient is assigned to an available pool, with probability equal to the fraction of idle servers in the system. For example, if pool i has two available servers and pool j has three available servers, the patient will be routed to any one of the five available servers with probability 1/5 (which is within pool i with probability 2/5 or within pool j with probability 3/5). The authors propose RMI as a routing algorithm that can easily be implemented in hospitals, possibly using patient ID numbers as sources of randomness. However, fairness does come at a cost of a higher probability of delay. Therefore, before implementation of RMI rather than a more efficient but less fair faster-servers-first (FSF) policy, it is important to determine if higher delays are acceptable.

One of the challenges of patient flow management from the ED to inpatient ward is that average patient LOS in the ED is typically measured in hours whereas the average LOS in an inpatient department is measured in days. This aspect of the problem was studied by Ramakrishnan et al. (2005) who were the first to propose a two-time-scale model for hospital patient flow from the ED to inpatient wards. Recently, Dai and Shi (2014) proposed another two-time-scale model for hospital patient flow management. They apply their model to assess the impact of inpatient discharge timing and bed capacity on ED boarding. The approach focuses on the changes that occur to the midnight census due to the medically required number of days in the patient's LOS along with the time of day that patients are admitted and discharged from the hospital. This work highlights the operational challenges in balancing inpatient admission and discharge timing.

2.9 Methodological Alternatives

In a review on ED modeling in the Emergency Medicine (EM) literature, (Wiler et al. 2011) refer to common approaches as formula-based, regression-based, time-series analysis, queueing theory, or discrete event simulation (DES). Formula-based approaches include ED crowding measures such as EDWIN. The authors note that regression-based and time-series forecasting methods are fairly well understood, but queueing theory and DES are not well understood and considered difficult to develop and use. However, in terms of ability to predict process improvement impact, queueing theory and DES are highlighted as better approaches. Therefore, the authors conclude that more collaboration between OM and EM researchers is needed in order to improve existing approaches.

In the studies reviewed so far, common methodologies include queueing, simulation, and optimization. As shown below in Table 2-2, most researchers use a single methodology, but some ED models incorporate multiple methodologies. For example, Zeltyn et al. (2011) use simulation, but adopt an MOL approach inspired by queueing theory. Some authors also create analytical queueing models based on a simplified process and test the robustness of their results with simulation (Almehdawe et al. 2013, Saghafian et al. 2012). Others use simulation optimization to incorporate the stochastic nature of the ED staff shift scheduling problem (Ahmed and Alkhamis 2009, Izady and Worthington 2012, Sinreich and Jabali 2007, Sinreich et al. 2012).

In most cases, the different methodology (denoted "Other" in Table 2-2) is forecasting or other statistical methods. Exceptions include the input-throughput-output framework (Asplin et al. 2003), the queueing network model of two EDs in a non-cooperative ambulance diversion game (Deo and Gurvich 2011), and the use of rough set theory, fuzzy measures, and cooperative game theory to help improve triage accuracy (Wilk et al. 2005). Other methodologies also include data mining (Ceglowski et al. 2006), gravity models with Bayesian updating (Congdon 2001), lean management (Holden 2011) and neural networks (Somoza and Somoza 1993).

Table 2–2: Methodological Alternatives

Study	Queueing	Simulation	Optimization	Other
Green et al. (2006)	X			
Green, Kolesar and Whitt (2007)	X			
Mayhew and Smith (2008)	X			
Cochran and Roche (2009)	X			
Mandelbaum et al. (2012)	X			
Allon et al. (2013)	X			
Deo and Gurvich (2011)	X			X
Zeltyn et al. (2011)	X	X		
Saghafian et al (2012)	X	X		
Almehdawe, Jewkes, and He (2013)	X	X		
Izady and Worthington (2012)	X	X	X	
Panayiotopoulos and Vassilacopoulos (1984)		X		
Saunders, Makens and Leblanc (1989)		X		
Lane, Monefeldt and Rosenhead (2000)		X		
Connelly and Bair (2004)		X		
Sinreich and Marmor (2005)		X		
Eldabi, Paul, and Young (2007)		X		
Fletcher et al. (2007)		X		
Hoot et al. (2008)		X		
Storrow et al. (2008)		X		
Fletcher and Worthington (2009)		X		
Khare et al. (2009)		X		
Bair et al. (2010)		X		
Paul, Reddy, and DeFlitch (2010)		X		
Wong et al. (2010)		X		
Klein and Reinhardt (2012)		X		
Abo-Hamad and Arisha (2013)		X		
Shi et al. (2015)		X		
Vassilacopoulos (1985)			X	
Beaulieu et al. (2000)			X	
Carter and Lapierre (2001)			X	
Ferrand et al. (2011)			X	
Sinreich and Jabali (2007)		X	X	
Ahmed and Alkhamis (2009)		X	X	
Sinreich, Jabali, and Dellaert (2012)		X	X	
Ceglowski et al. (2007)		X		X
Tierney et al. (1986)				X
Graff et al. (1993)				X
Somoza and Somoza (1993)				X
Congdon (2001)				X
Jones, Joy, and Pearson (2002)				X
Asplin et al. (2003)				X
Schull et al. (2004)				X
Ramakrishnan et al. (2005)				X
Wilk et al. (2005)				X
Asplin, Flottemesch and Gordon (2006)				X
Hoot et al. (2007)				X
Schull, Kiss, and Szalai (2007)				X
Holden (2011)				X
Soremekun, Terwiesch, and Pines (2011)				X
Wiler, Griffey, and Olsen (2011)				X
Powell et al. (2012)				X
Dai and Shi (2014)				X

Considering the complexity of the ED patient flow process, it is not surprising that simulation is the most commonly used methodology in Table 2-2. Additional ED-specific examples are also identified in a systematic review of simulation projects investigating ED crowding from 1970 to 2006 (Paul et al. 2010). A key point both from that ED specific review and healthcare simulation modeling in general (Eldabi et al. 2007) is that most research incorporates only one part of the healthcare system such as an ED, without consideration of the complex interconnections that exist among healthcare system components. It is also important to note that there are a variety of simulation techniques available and while discrete event simulation (DES) is often used for micro-level analysis of a specific department, system dynamics (SD) simulation is used for aggregate analysis of the interaction between components. In this review, the main example of SD simulation is in modeling the interface between the ED and inpatient units for admitted patients who remain boarding in ED hallways (Lane et al. 2000, Wong et al. 2010).

DES models are traditionally developed using specialized simulation software programs such as Arena, MedModel, or Simul8. However, Klein and Reinhardt (2012) have shown that the complexity of the ED patient flow process can be modeled with spreadsheet simulation. Developed using Microsoft Excel 2007 (Microsoft, Redmond, WA), the spreadsheet simulation was compared to the Khare et al. (2009) MedModel simulation and a statistical equivalence test formally demonstrated that spreadsheet simulation is equally effective. Spreadsheet software is easy to use and widely available, at a fraction of the cost of DES software. Coding spreadsheet simulations may be more challenging since it requires a different and more novel expertise than traditional computer programming. However, spreadsheets can be organized to reference existing datasets thus minimizing the burden of copying and likelihood of transcription errors and information leakage. Output analysis can also be customized with user-specific performance statistics and charts.

For a two-week horizon, the spreadsheet simulation model reserves a set of rows where each row represents a minute: $1, 2, 3, ..., 21,600$ where that last number represents fifteen

days worth of minutes $[(1 + 14) * 24 * 60]$ which includes an initial warm-up period of 1,440 minutes ($= 24 * 60 =$ one day). From a modeling perspective, the key insight comes from understanding the difference in how Excel processes a simulation rather than a more traditional program such as MedModel. In Excel, the entire system state is summarized in a *minute* sequence, where each rows is based on its own random number results along with its preceding minute's values. Performance statistics are derived based on the time series of the states of the system at each minute. In MedModel, the focus is on the stay of a *patient* through the ED, from arrival to admission or discharge, and statistics are tallied and updated based on the performance of the ED resources for that patient.

2.10 Emergency Care in Rural Areas

In this systematic review of the literature, we have provided a detailed account of the state of Operations Research for Emergency Care. There have been many studies completed on this topic, however, research on emergency care has only considered the urban context. Surprisingly, none of the studies included in this comprehensive review investigated Emergency Care in Rural Areas. In urban areas, studies have been completed to consider some of the unique features of ED patient flow including triage and lab tests, but urban ED patient flow research is not exhaustive. In particular, we also identified from our review that the ED consultation process has not been studied. In the next chapter, we study the ED consultation process by considering the workflow decisions of specialists. ED consultations and inpatient ward care are handled by specialists, so we are very interested in Specialist Care. In both of the next two chapters, we examine understudied healthcare challenges for specialists in rural areas.

Chapter 3
Specialist Care in Rural Hospitals: From Emergency Department Consultation to Inpatient Ward Discharge

3.1 Introduction

This chapter contributes to the literature on Emergency Department (ED) outflow to inpatient wards. To the best of our knowledge, this is the first study examining ED crowding from the specialist's perspective. The multi-departmental process from the ED to inpatient ward begins when specialists receive ED consultation requests. Previous models of inpatient operations do not consider the importance of ED consultations, focusing instead on *bed-based capacity management* considering only patient flow after admission has been confirmed. Furthermore, while most research is conducted in *urban* hospitals, we study the additional challenges faced by specialists in *rural* hospitals.

We propose a stochastic dynamic programming framework for specialist care that includes a *Single Role Model* and a *Dual Role Model*. Should specialists give priority to ED consultation requests or give priority to inpatient discharges? Defining *C-Priority* as the Single Role Model policy where specialists always give priority to ED consultations and *D-Priority* as the policy where specialists always give priority to inpatient discharges, we can compare the effectiveness of these and other decision making policies to the optimal policy. We apply the proposed modeling framework to data sets developed from two corresponding case studies. The traditional rural case is more complex since Internal Medicine Specialists (Internists) take on a dual role as the Intensive Care Unit (ICU) physician and Internist on call for ED consultations and inpatient care in the medicine wards.

In the remainder of the chapter, we review the most relevant literature in section 3.2, define our stochastic dynamic programming models in section 3.3 and describe the structure of optimal policies in section 3.4. We then describe our study setting and data sources in

section 3.5, and report results of the application of our Single and Dual Role Models to two study hospitals in section 3.6. A discussion is provided in section 3.7 and we summarize our findings with concluding remarks in section 3.8.

3.2 Literature Review

One of the main ED patient flow bottlenecks is getting admitted patients out of the ED. The medical community have been reporting this *ED boarding problem* with calls for system wide solutions to manage the challenges of patient populations with increasing severity of illness (Forster et al. 2003, Trzeciak and Rivers 2003, Falvo et al. 2007). At Northwestern Memorial Hospital in Chicago, Khare et al. (2009) used simulation to study the effectiveness of ED patient flow strategies on throughput and output. Their analysis demonstrated that increasing the number of ED beds did not reduce length of stay (LOS) but increasing the rate that admitted patients depart the ED did result in a significant reduction in LOS. In order to investigate potential solutions to reduce boarding times, models have been developed to test smoothing inpatient discharge timing by day of week (Wong et al. 2010) or by time of day (Powell et al. 2012, Shi et al. 2015). Along with early inpatient discharge policies, new institutional guidelines are putting pressure on specialists to provide faster ED consultation response times. In a Pennsylvania study of the impact of an institutional guideline for timely ED consultations, Geskey et al. (2013) found that reduced ED consultation response times were obtained after implementation of the guideline. However, the guideline also resulted in longer inpatient discharge times, leaving unanswered questions on how to best manage the timing of specialist care considering both ED consultations and inpatient discharges.

One of the challenges of patient flow management from the ED to inpatient ward is that average patient LOS in the ED is typically measured in hours whereas the average LOS in an inpatient department is measured in days. This aspect of the problem was studied by Ramakrishnan et al. (2005) who were the first to propose a two-time-scale model for hospital patient flow from the ED to inpatient wards. Recently, Dai and Shi (2014) proposed another two-time-scale model for hospital patient flow management. They apply

their model to assess the impact of inpatient discharge timing and bed capacity on ED boarding. The approach focuses on the changes that occur to the midnight census due to the medically required number of days in the patient's LOS along with the time of day that patients are admitted and discharged from the hospital. This work highlights the operational challenges in balancing inpatient admission and discharge timing. Additional bed-based capacity management approaches include Mandelbaum et al. (2012) and Allon et al. (2013). Based on a large Israeli hospital, Mandelbaum et al. (2012) propose an inverted-V queueing system with a single centralized queue and multiple server pools: the pools represent the wards and servers are beds. Allon et al. (2013) propose a two-station queueing network model with the first station as the ED and the second station as the inpatient department. Using heavy traffic approximations applied to hospitals in California, the study findings highlight the fact that ED crowding is a multi-departmental problem: the extent that a hospital goes on ambulance diversion varies on the size and occupancy of both the ED and inpatient departments.

The relevant healthcare operations management literature also includes studies that model healthcare providers as servers. Although these studies do not necessarily involve inpatient operations, they are equally relevant to our work. Green et al. (2006a) developed a discrete time finite horizon dynamic programming framework to model the decision making challenges for patient selection in diagnostic medical facilities. Over the course of the day, these services are provided to scheduled outpatients and non-scheduled inpatients as well as emergency patients from the ED. ED patients are given priority in the model, and the main challenge is to determine who to serve next when there are both inpatients and outpatients waiting for diagnostic services. The modeled decision making challenge incorporates a probability of no-shows for scheduled outpatients and the uncertainty in the time of arrivals of emergency and inpatient requests for diagnostic services. Optimal policies and heuristics are developed to evaluate various capacity management strategies including threshold appointment schedules where service slots before a threshold time are used for scheduled

appointments with later slots left open. The optimal policies indicate that the decision on which patient to serve next depends on the time of the day and the number of outpatients and inpatients waiting for diagnostic services.

Yankovic and Green (2011) proposed a queueing system with two types of servers: beds and nurses. With this model, the authors show how admissions and discharges may be blocked when inpatient beds are occupied or due to a lack of nurse availability. The system is modeled as a quasi-birth and death (QBD) process which is then solved with matrix analytical methods. Performances measures derived for the system include the probability of delay for a bed or a nurse, bed utilization, nurse utilization, and the probability that discharges are blocked.

We are not the first to study the workflow decisions of physicians. Dobson et al. (2013) study the workflow decisions of ED physicians and consider the impact of interruptions on physician workload. ED physicians perform an initial assessment and order lab tests for new patients after triage. Once lab tests are returned, ED physicians also need to re-assess and determine a treatment plan. Faced with an overloaded system of new patients waiting at triage and existing patients waiting for re-assessment, ED physicians choose which patient to see next. If an ED physician decides to serve new patients after triage, it will reduce the door-to-doctor time for the new patient. However, that decision will also increase the LOS for the existing patient whose ED care may have been completed sooner if seen first. While the model does not consider the role of specialists in the admission process and other complexities of ED patient flow, the analysis does highlight the important trade-off in serving new or existing ED patients. The model also considers aspects rarely included in other ED patient flow models such as patient reneging for those who leave without being seen (LWBS) and the impact of interruptions on physician workload. Recently, Huang et al. (2015) also modeled workflow decisions of ED physicians. They propose decision making policies regarding when to serve new triage patients who have not yet been seen and when to serve in process patients who require re-assessment after lab tests or other investigation

has been completed. Again, while not all complexity of the ED service process is considered, the authors do consider multiple patient classes with time deadlines for new patients after triage and the feedback component of ED patient flows where patients may require service multiple times during their ED LOS before the disposition decision to discharge from the ED or admit the patient to a speciality ward for further hospital care.

However, to the best of our knowledge we are the first to study the workflow decisions of specialists. In previous studies, the admission process is considered to start with an admission request. However, in studying the admission process from the specialist's perspective, it is clear that the admission process really begins when a specialist receives an ED consultation request. Specialist consultation is an important part of ED patient flow which has received little research attention. We study the role of specialists in patient flow management from the ED to inpatient wards, and consider the impact of ED consultation, inpatient care, and inpatient discharge timing.

3.3 Stochastic Dynamic Programming Models

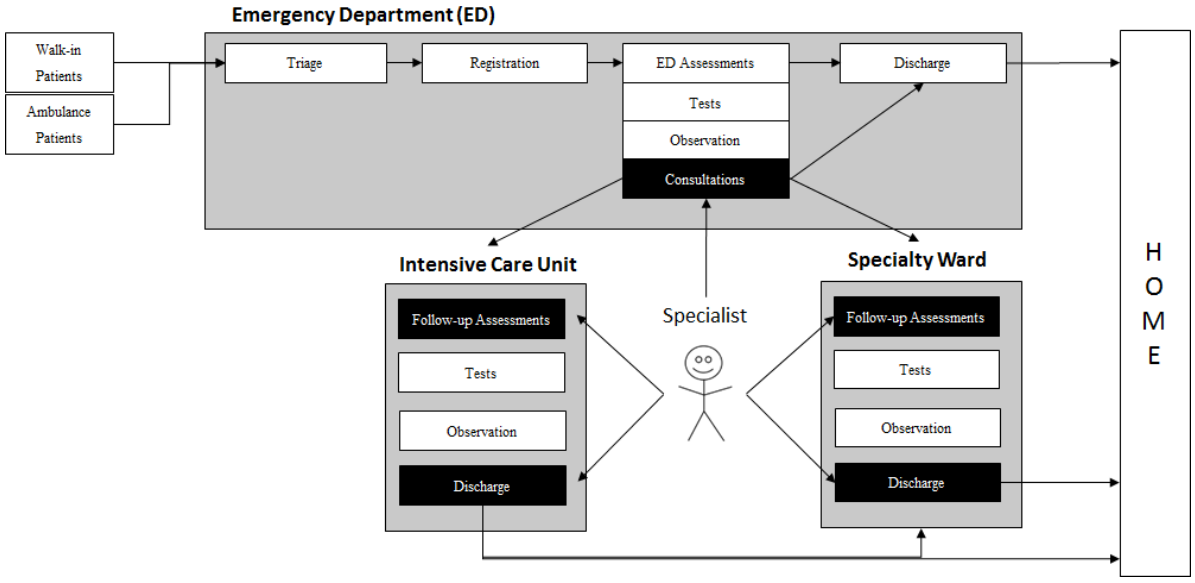

Figure 3–1: Specialist Workload in the Hospital

Specialists work on call in the hospital, with duties spanning multiple departments as shown in Figure 3-1. In some cases, patients who arrive to the ED can be treated and

released under the care of an ED physician. After triage and registration, the ED physician and nurse complete their assessments. The ED physician may investigate by requesting one or several lab tests (i.e. blood/fluids or imaging) or keep the patient under observation. When uncertainty or complexity arises in diagnosis or treatment, ED physicians call on specialists to come to the ED for consultations. When a specialist goes to the ED for consultation, the specialist consultation includes an admission order or a discharge order from the specialist. For patients requiring further treatment, specialists take over care from the ED physician once admitted to the inpatient ward. Meanwhile, specialists need to provide follow-up care to patients in specialty wards and decide when inpatients are ready for discharge. In rural hospitals, specialists are responsible for the ICU, providing follow-up care and admission and discharge decisions for ICU patients as well. While specialists also offer outpatient clinics and perform specialty procedures including surgeries, such work is typically scheduled separately from the on call duties depicted in Figure 3-1.

In some rural hospitals including one of our study hospitals, ICU physicians and Internists are staffed at the same time, each taking on a *single role*. However, the traditional situation for rural hospitals involves one specialist taking on both roles at the same time, serving as the ICU physician and Internist on call. This *dual role* requires the specialist to consider additional work, and as a result, our corresponding Dual Role Model is more complicated than our Single Role Model. While our focus in this study is on rural hospitals, our proposed Single and Dual Role Models would also be appropriate for urban hospitals that have practices similar to either of our study hospitals. We begin with a description of the Single Role Model and then describe the Dual Role Model as an extension to the Single Role Model.

3.3.1 Single Role Model

Consider a one day model with the specialist's time on a given day divided into N discrete time slots (periods). The state of the system is represented by (c, b, d) where c represents the number of patients waiting in the ED for specialist consultation, b represents

the number of patients boarding in the ED for admission (ED boarders), and d represents the number of patients in the hospital ward waiting for their last care visit from a specialist. One of our modeling goals is to minimize the number of state variables so that the model can be solved more efficiently. One novel element in our model is that we do not need an additional variable to keep track of inpatient bed capacity. When the hospital ward is full with no inpatient beds available, b has a positive value for the number of ED boarders. When there is available inpatient bed capacity, we use the same b variable with a negative value to represent inpatient bed capacity. In that case, -b represents the number of available inpatient beds in the hospital ward. In the case of a full hospital ward with all inpatient beds filled but no ED boarders, the value of b is zero. The patients included in the d variable are those who are expected to be discharged from the hospital ward today. This variable is determined at the beginning of the horizon, possibly during morning rounds. These patients will be ready for discharge after the specialist completes their final care visit and issues a discharge order. Note that our state space excludes the number of patients waiting for follow-up specialist care for those patients who we already know will remain in the hospital ward overnight. We assume that these follow-ups will be handled by the specialist when the state space is $(0, b, 0)$. That is to say, ED consultations and inpatient ward discharges are given priority over regular follow-up hospital ward care.

Let $\gamma_i = C, D$ represent the action taken by the specialist in period i where C = Emergency Department Consultation and D = Hospital Ward Care (for Discharge). The D action represents the last care visit required before discharge for a patient in the hospital ward. We also define w_c, w_b, and w_d to represent the per period waiting costs for patients waiting for ED consultation, boarding for admission, and waiting for discharge, respectively. At the end of the day, we also define an end of horizon penalty cost function denoted $g(c, b, d)$.

In this model, there is one on call specialist available to handle the C and D actions in each period. We assume that the C and D actions take the same amount of time, occupying

the specialist for one period. New consult requests are assumed to arrive before the specialist's decision at the beginning of each period. Our models are most appropriate for inpatient departments such as Internal Medicine which receives almost all admissions from the ED. Therefore, we assume that there are no elective admissions and the only input source is from the ED. We assume, as observed in practice, that high acuity ED consultations have the highest priority. We handle this situation without requiring an additional variable in our state space. While urgent requests are of great importance, from a decision making point of you, the decision is trivial. When there is an urgent request from the ED, no matter what the state of the system, the specialist's action will be C. While boarding is very common when waiting for admission to medicine wards, boarding for admission to ICU is much less common and ICU bumping (Dobson et al. 2010) is assumed to ensure that ICU capacity is always available for high acuity admissions. The final model assumption is that an inpatient bed is available upon discharge for use in the next period. In practice, there may be longer bed turnaround times due to cleaning, or other delays if patients need to wait for pick up or if a patient is waiting for an alternate level of care (ALC) ward downstream which does not have space available.

For the arrival process, we denote the arrival probability for ED consultation requests by λ_c and let $(1 - z)$ represent the fraction of high acuity ED consultation requests. We assume that all high acuity ED consultation requests result in admission to ICU and for the remaining consultation requests, denote p_a as the admission probability to the medicine wards. We can now define the Single Role Model with the following stochastic dynamic program:

$$V^* = \min_{\gamma}[V_1^{\gamma_1}(c,b,d)] \tag{3.1}$$

$$\begin{aligned} V_i^{\gamma_i}(c,b,d) &= cw_c + max(0,b)\cdot bw_b + dw_d \\ &\quad +(1-z)\lambda_c[V_{i+1}^{\gamma_{i+1}}(c,b,d)] + z\lambda_c[H_{i+1}^{\gamma_{i+1}}(c+1,b,d)] \\ &\quad +(1-\lambda_c)[H_{i+1}^{\gamma_{i+1}}(c,b,d)], i = 1,...,N. \end{aligned} \tag{3.2}$$

$$H_i^{\gamma_i}(c,b,d) = \begin{cases} min\begin{pmatrix} p_a(V_i^{\gamma_i}(c-1,b+1,d)) \\ +(1-p_a)V_i^{\gamma_i}(c-1,b,d), \\ V_i^{\gamma_i}(c,b-1,d-1) \end{pmatrix} & \text{if } c \geq 1,\ d \geq 1, \\ p_a(V_i^{\gamma_i}(c-1,b+1,0)) \\ +(1-p_a)V_i^{\gamma_i}(c-1,b,0) & \text{if } c \geq 1,\ d = 0, \\ V_i^{\gamma_i}(c,b-1,d-1) & \text{if } c = 0,\ d \geq 1, \\ V_i^{\gamma_i}(0,b,0) & \text{if } c = d = 0 \end{cases} \tag{3.3}$$

$$H_N^{\gamma_N}(c,b,d) = V_N^{\gamma_N}(c,b,d) = g(c,b,d) \tag{3.4}$$

The objective is to minimize the total expected daily waiting costs. For a given period i, the waiting costs are the sum of the first three terms in (3.2). The second term accounts for the situation when b is negative (i.e. no ED boarders possibly with available inpatient bed capacity). In that case, whenever $b \leq 0$, the second term is zero, and we ensure that the waiting cost is nonzero only if there are boarders. The possible transitions to and from state (c,b,d) included in (3.2),(3.3) are illustrated in Figure 3-2.

In period i, if there is a new ED consultation request and the specialist's action is D, the process will transition to state $(c+1,b-1,d-1)$ for the next period. Similarily, if the specialist's action is D but there is no new ED consultation request, the next state will be $(c,b-1,d-1)$. If there is no new ED consultation request and the specialist's acton is C, the process will transition to state $(c-1,b+1,d)$ with probability p_a or to state $(c-1,b,d)$ with probability $(1-p_a)$. Similarily, if there is a new ED consultation request and the specialist's action is C, the process will transition to state $(c,b+1,d)$ with probability p_a or

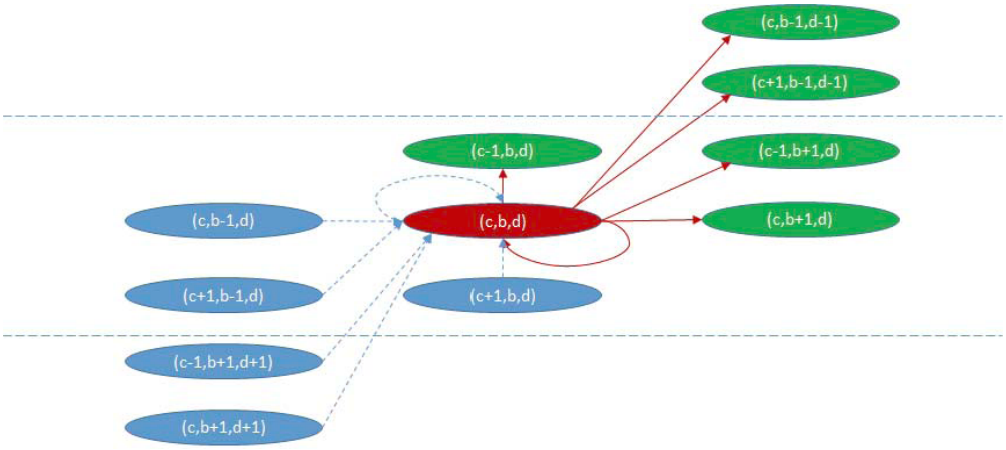

Figure 3–2: State Transitions for the Single Role Model

remain in state (c, b, d) with probability $(1 - p_a)$. Note in Figure 3-2 that there are the same number of transitions in to node (c, b, d) as there are from node (c, b, d), so the state space will not explode. The boundary condition (3.4) includes the end of day penalty function which depends on the state of the system (c, b, d) at the end of the day.

3.3.2 Dual Role Model

Next, we describe the Dual Role Model, applicable to rural hospitals that have one Internist staffed to handle both the role of ICU physician and Internist on call at the same time. Here we extend the model described in the previous section, highlighting the additional variables required to handle the additional complexity of the dual role.

With appropriate assumptions as observed in practice, we identified a state space with only one additional variable, denoted f. Therefore, the state of the system is represented by (c, b, f, d) where f represents the number of patients in the ICU waiting for follow-up visit from a specialist. We also update the action space to reflect the work required for follow-up specialist care in the ICU, denoted F. Let $\gamma_i = C, F, D$ represent the action taken by the specialist in period i where C = Emergency Department Consultation, F = ICU Follow-up Care and D = Hospital Ward Care (for Discharge). Finally, we add an additional waiting cost variable denoted w_f to represent the per period waiting costs for patients waiting for

ICU follow-up care. At the end of the day, we also update the end of horizon penalty cost function to $g(c, b, f, d)$. We can now define the Dual Role Model as follows:

$$V^* = \min_{\gamma}[V_1^{\gamma_1}(c, b, f, d)] \tag{3.5}$$

$$\begin{aligned} V_i^{\gamma_i}(c, b, f, d) = \; & cw_c + max(0, b) \cdot bw_b + fw_f + dw_d \\ & +(1 - z)\lambda_c[V_{i+1}^{\gamma_{i+1}}(c, b, f, d)] + z\lambda_c[H_{i+1}^{\gamma_{i+1}}(c + 1, b, f, d)] \\ & +(1 - \lambda_c)[H_{i+1}^{\gamma_{i+1}}(c, b, f, d)], i = 1, ..., N. \end{aligned} \tag{3.6}$$

$$H_i^{\gamma_i}(c, b, f, d) = \begin{cases} min \begin{pmatrix} p_a(V_i^{\gamma_i}(c - 1, b + 1, f, d)) \\ +(1 - p_a)V_i^{\gamma_i}(c - 1, b, f, d), \\ V_i^{\gamma_i}(c, b, f - 1, d), \\ V_i^{\gamma_i}(c, b - 1, f, d - 1) \end{pmatrix} & \text{if } c \geq 1,\ f \geq 1,\ d \geq 1, \\ min \begin{pmatrix} p_a(V_i^{\gamma_i}(c - 1, b + 1, f, d)) \\ +(1 - p_a)V_i^{\gamma_i}(c - 1, b, f, d), \\ V_i^{\gamma_i}(c, b - 1, f, d - 1) \end{pmatrix} & \text{if } c \geq 1,\ f = 0,\ d \geq 1, \\ min \begin{pmatrix} p_a(V_i^{\gamma_i}(c - 1, b + 1, f, d)) \\ +(1 - p_a)V_i^{\gamma_i}(c - 1, b, f, d), \\ V_i^{\gamma_i}(c, b, f - 1, d) \end{pmatrix} & \text{if } c \geq 1,\ f \geq 1,\ d = 0, \\ min \begin{pmatrix} V_i^{\gamma_i}(c, b - 1, f, d - 1), \\ V_i^{\gamma_i}(c, b, f - 1, d) \end{pmatrix} & \text{if } c = 0,\ f \geq 1,\ d \geq 1, \\ p_a(V_i^{\gamma_i}(c - 1, b + 1, 0, 0)) \\ +(1 - p_a)V_i^{\gamma_i}(c - 1, b, 0, 0) & \text{if } c \geq 1,\ f = d = 0, \\ V_i^{\gamma_i}(0, b, f - 1, 0) & \text{if } c = 0,\ f \geq 1,\ d = 0, \\ V_i^{\gamma_i}(0, b - 1, 0, d - 1) & \text{if } c = f = 0,\ d \geq 1, \\ V_i^{\gamma_i}(0, b, 0, 0) & \text{if } c = f = d = 0 \end{cases} \tag{3.7}$$

$$H_N^{\gamma_N}(c, b, f, d) = V_N^{\gamma_N}(c, b, f, d) = g(c, b, f, d) \tag{3.8}$$

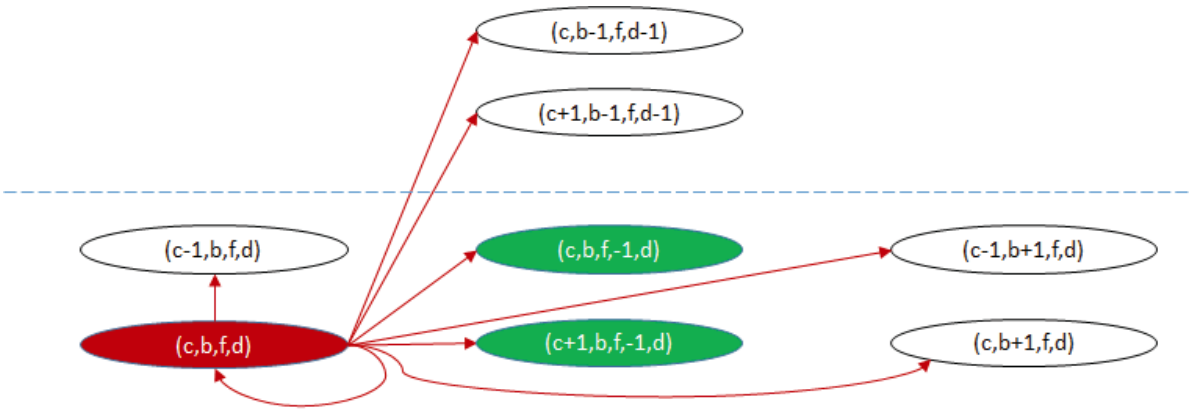

Figure 3–3: State Transitions for the Dual Role Model

.

As in the Single Role Model, the objective is to minimize the total expected daily waiting costs. For a given period i, the waiting costs are the sum of the first *four* terms in (3.6). The possible transitions to and from state (c, b, f, d) included in (3.6),(3.7) are illustrated in Figure 3-3.

The transitions where f does not change are analagous to those in the Single Role Model. There are two additional transitions when the specialist's action is F. In period i, if there is a new ED consultation request and the specialist's action is F, the process will transition to state $(c+1, b, f-1, d)$ for the next period. Similarily, if the specialist's action is F but there is no new ED consultation request, the next state will be $(c, b, f-1, d)$. Note that we only have transitions for decrementing the value of f, similar to d. We originally thought that we would need to extend the state space to account for new arrivals of ICU patients to the ICU and consider state transitions where f is incremented. However, this is not necessary since the specialist will see the new ICU patients in the ED, which we cover in our model as high acuity ED consultation requests. Typically, the ICU nursing staff takes care of the patient in the ICU after admission. The ICU staff follows the specialist's documented treatment plan and calls the specialist if clarification is required. New ICU patients stay in the hospital for several days, so the specialist's workload for the *next* day will include follow-ups for the ICU patients admitted on the previous day, which is included in the initial value of f established

at morning rounds. We also update the boundary condition to consider f, so that (3.8) includes the end of day penalty function which depends on the state of the system (c, b, f, d) at the end of the day.

3.4 Optimal Policy Structure

Consider the Single Role Model with waiting cost parameters such that $w_c > w_b > w_d$ and an end of day penalty cost function $g(c, b, d) = cy_c + by_b + dy_d$ such that $y_d > y_b > y_c$. That is to say, during the day per period waiting costs are highest for patients waiting for ED consultation and lowest for patients waiting for discharge in the inpatient ward. However, overnight waiting costs are higher for patients who need to stay an extra night in the ward, and lower for patients who stay overnight in the ED. As described in section 5, these parameters reflect that an ED physician is available overnight but there is no doctor staffed overnight in the medicine wards.

Our optimal policy then has the following form:

$$\gamma_i = \begin{cases} C & \text{if} \quad c > 0, d = 0, \\ & \text{or} \quad c > 0, d > 0, i \le i^*, b \le b^X, \\ & \text{or} \quad c > 0, d > 0, i^* \le i \le j^*, b \le b^Y, \\ & \text{or} \quad c > 0, 0 < d \le d_i^*, i^* \le i \le j^*, b^Y \le b \le b^X, \\ & \text{or} \quad c > 0, 0 < d \le d_i^*, i > j^*, b^Y \le b \le b^X, \\ D & \text{if} \quad c > 0, d > 0, i^* \le i \le j^*, b > b^X, \\ & \text{or} \quad c > 0, d > d_i^*, i^* \le i \le j^*, b^Y \le b \le b^X, \\ & \text{or} \quad c > 0, d > d_i^*, i > j^*, b^Y \le b \le b^X, \\ & \text{or} \quad c > 0, d > 0, b > b^X, \\ & \text{or} \quad c = 0, d > 0 \end{cases} \tag{3.9}$$

Figure 3-4 illustrates an example of the optimal policy for the N-period discrete time finite horizon model when the patient waiting costs per period for ED consultation, boarding, and discharges are 5, 3, and 1 respectively. In this example, the initial state is (0,0,15) and the end of day (overnight) penalty costs for patients waiting for ED consultation, boarding for admission, or waiting for discharge are 10, 20, and 40 respectively. This example also has the expected total number of new ED consultation requests set to 15 for the day and an admission probability of 0.7 after ED consultation.

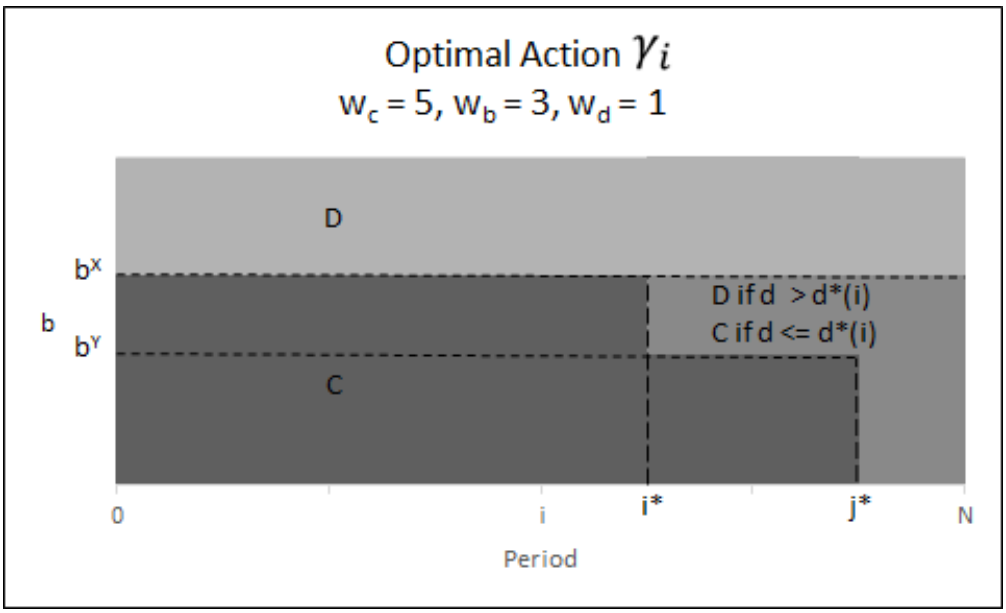

Figure 3–4: Optimal Policy Structure: Single Role Model

The optimal action depends on the period number i and the state variables (c, b, d). We observe that the optimal policy has different forms throughout the day (horizon), with boarding thresholds and end-of-horizon effects. At the beginning of the horizon up to i^*, the optimal policy is mainly based on a boarding threshold (i.e. D if $d > 0$ and $b > b^X$; C if $c > 0$ and $b \leq b^X$). As the end-of-horizon approaches, the policy changes such that 1) the boarding threshold is reduced and 2) the end-of-horizon effect also matters. We then have an optimal policy of the form: D if $d > 0$ and $(b > b^X$ or $(b > b^Y$ and $d > d_i^*)$; C if $c > 0$ and $b <= b^Y$ or $(b <= b^X$ and $d <= d_i^*)$. From period j^* to the end of the day, the end-of-horizon effect becomes more prominent and the optimal policy takes on the form:D if $d > 0$ and $(b > b^X$ or $d > d_i^*$; C if $c > 0$ and $(b <= b^X$ and $d <= d_i^*)$. Typically, $d_i^* = N - i$. As a result, the optimal policy will usually avoid the end-of-day penalty costs and ensure that those patients are discharged before the end of the horizon.

3.5 Study Setting and Data Sources

We apply our models to two regional hospitals in Nova Scotia, Canada. Each region also has two additional community hospitals staffed with ED physicians but Internal Medicine inpatient care is not provided in the community hospitals. When inpatients at community

hospitals require Internal Medicine consultation or Intensive Care, they are transferred to the regional hospital by ambulance.

In both regional hospitals, Internists work on call for ED consultations, provide inpatient care in the medicine wards, and also provide Intensive care in the ICU. In South Shore Regional Hospital (SSRH), the traditional rural approach is used with one Internist on call for ICU, ED consultations and inpatient care in the medicine wards. SSRH Internists cover on call day shifts from 8am-5pm and/or on call night shifts from 5pm-8am. Due to ED crowding, in Yarmouth Regional Hospital (YRH), management decided to staff a separate Internist for the ICU similar to an urban hospital. YRH staffs Internists on call for 24 hour shifts with one Internist on call for ICU, and another Internist generally works on call to handle ED consultations and inpatient care in the medicine wards.

We observed the work of *all* Internists on call in the two regional hospitals. Decision Support Analysts extracted data from hospital information systems. During the period from November 1, 2014 to June 30, 2015, SSRH had 12062 ED visits and YRH had 17178 ED visits.

Specialist consultation and service times are not recorded in hospital information systems, so we worked with the two physician groups to collect Internist data including service times for ED consultations and inpatient follow-up care. Before observing Internists, we thought that specialists would have a setup time to walk to the ED and back to the wards which could lead to batching consultation requests. However, we timed the setup time while observing Internists and discovered that the time was at most two minutes. While we did not observe this as an issue in our study hospitals, if the setup time is considerably long in other hospitals, then the service time could be adjusted to incorporate the setup time. After observing Internists, it also became clear that service time is much more than a patient visit. Indirect care operations are often overlooked, but we determined that both direct and indirect care operations must be considered to accurately reflect specialist workload. Therefore, service times include reviewing patient charts, x-rays and medical history, patient

visits, updating progress notes, written orders, prescriptions, tests, procedures, recording the decision to admit, follow-up / re-assess or discharge, and dictation on the phone to complete the assessment. It is also worth noting that the discharge process is more than just signing-off. Internists need to visit each patient and complete an assessment to ensure that they are ready to go home. The Internist also updates progress notes, checks to ensure that all prescriptions are appropriate before completing written orders and dictation by phone to complete the assessment.

Table 3–1: Single Role Model Parameters

Description	Variable	YRH	SSRH
Number of Periods per day	N	20	20
Fraction of High Acuity ED Consult Requests	1-z	0.22	0.33
Admission Probability to Medicine Wards	p_a	0.38	0.42
Waiting Costs (base case)	(w_c, w_b, w_d)	(3,2,1)	(3,2,1)
End of day Penalty Costs	$g(c, b, d)$	$10c + 20b + 40d$	$10c + 20b + 40d$

Parameters for the Single Role Model are shown in Table 3-1. Each day has 20 time slots: 45 minute periods from 9:00AM to 11:59PM. We selected the 45 minute time slots based on what we observed in practice. Our Internist data from YRH include 538 patients visits, including 63 ED consultations resulting in admission to medicine wards and 101 ED consultations resulting in discharges from the ED. The mean (standard deviation) for those ED consultations are 46.93 minutes (16.60 minutes) and 39.54 minutes (16.78 minutes) respectively. For inpatient discharges from the medicine wards, our Internist data includes 54 observations with mean (standard deviation) of 27.55 minutes (17.06 minutes). While discharges from the medicine wards do take less time than ED consultations, inpatient discharges from the medicine wards take much longer than regular follow-ups in the medicine wards. In 137 observations, regular follow-ups occupied the specialist for a mean (standard deviation) of only 12.36 minutes (7.90 minutes).

We also use our Internist data to determine the values of two important model parameters: the fraction of high acuity ED consultation requests (1-z) and the admission probability

to medicine wards (p_a). Recall that no matter what the state of the system, whenever there is a high acuity ED consultation request, the specialist's next action will be C. Those urgent consultation requests typically result in admission to the ICU. On the other hand, the remaining regular ED consultation requests require admission to the medicine wards with probability p_a. The rest of the ED consultation requests result in discharge from the ED with probability $1 - p_a$.

Hospital information systems data is used to determine the other parameters required for the model. In each day with the initial state of the system (c, b, d), c includes new consult requests from 12:00AM to 8:59AM. While the initial value of b depends on the actions of the previous day, the initial value of the d variable is established from hospital discharge data for each day. In practice, the specialist knows the value of the d variable at period 0 since this is the number of patients that are ready for discharge on a given day. The specialist knows this information either from taking care of inpatients on previous days or it is determined during morning rounds. Recall that the d variable is only decremented throughout the day since new inpatient admissions will have a LOS greater than one day.

The arrival rate of new consultation requests is λ_c, taken from hospital data with a different rate depending on the day. We assume that consultation requests follow a Poisson process. While we did observe time-varying arrivals to the ED by time of day, the requests for specialist consultation do not follow the same pattern. Figure 3-5 shows YRH Internist data collected on ED consultation requests by hour of day for 215 ED consultation requests from 65 different days. After accounting for the overnight batch of consultation requests captured in our inital c variable at period 0, the remaining ED consultation requests are relatively homogeneous throughout the day.

The model parameters are the same for the Dual Role Model with the addition of the f variable for ICU follow-up care requirements. Our Internist data includes 93 regular ICU follow-ups and 44 ICU follow-ups resulting in discharge. The mean (standard deviation) for those ICU follow-ups are 21.87 minutes (10.94 minutes) and 17.68 minutes (5.65 minutes)

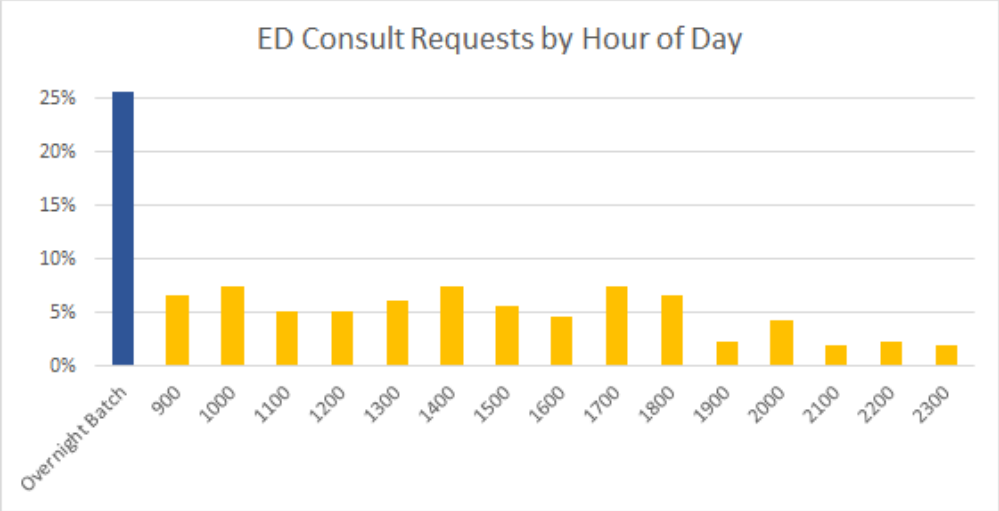

Figure 3–5: YRH - ED Consultation Requests by Hour of Day

respectively. As a result, we assume that an ICU follow-up occupies the specialist for half of a 45 minute period. Suppose that at morning rounds, the specialist determines that there are four patients waiting for follow-up in the ICU. In that case, it would take two 45 minute periods for the specialist to complete all four ICU follow-ups. We use hospital information systems data to determine the number of ICU follow-ups required for each of the 242 days.

In our base case, the waiting costs are $(w_c, w_b, w_d) = (3, 2, 1)$ for the Single Role Model and $(w_c, w_b, w_f, w_d) = (3, 2, 5, 1)$ for the Dual Role Model. These parameters are not available through the data from the two study hospitals. For our purposes, it is the relative magnitude of these variables in comparison with each other that is more important than the precise values for these parameters. The relative values were set subjectively in consultation with the physician coauthors based on their conceptions of medical priorities.

We take a patient health perspective in establishing our base case parameters. The sickest patients with applicable daytime waiting costs are those waiting for ICU follow-up and we give this group the highest waiting cost. Note that high acuity ED consultations are excluded here since those patients are always given the highest priority and do not accumulate waiting costs. For the remaining patients, during the day, per period waiting costs are likely highest for patients waiting for ED consultation and lowest for patients waiting for discharge

in the inpatient ward. However, overnight waiting costs are likely higher for patients who need to stay an extra night in the ward, and lower for patients who stay overnight in the ED. From a patient health perspective, it is reasonable to assume that patients waiting in the ED for specialist consultation during the day will have a high waiting cost. It is also reasonable to assume that waiting cost is lower when boarding since the patient has received an assessment from a specialist during their ED consultation and the requirement for admission has been confirmed. Meanwhile patients in the ward waiting for discharge likely have the lowest daytime waiting cost as they have had several days of care and treatment with improved health. On the other hand, waiting costs could be different overnight. Considering that the ED is staffed through the night with an emergency physician, an ED patient can be seen by the emergency physician whenever the need arises. If the patient has trouble sleeping, an ED patient can have a prescription for sleeping pills, for example. However, in an inpatient ward, typically there is no doctor staffed during the night, so the overnight waiting cost for these patients is highest, and those who have trouble sleeping need to wait until morning to see a doctor. ED boarders have access to an emergency physician overnight, but boarding in a hallway is less comfortable than waiting in a private ED treatment space, so waiting cost is likely higher for boarders than other ED patients.

In the case studies that follow in the next section, we include a parametric analysis on the waiting cost parameters: $w_c = 3, 4, 5$; $w_b = 1, 2, 3$; and $w_d = 1, 2, 3$. The end of day penalty costs are $g(c, b, d) = 10c + 20b + 40d$ for the Single Role Model and $g(c, b, f, d) = 10c + 20b + 80f + 40d$ for the Dual Role Model.

3.6 Case Studies

Our Single Role Model represents the current practice of YRH and the Dual Role Model represents the current practice of SSRH. We apply the Single Role and Dual Role Models to both hospitals and analyze the effectiveness of different policies with each model. Under the Single Role Model, we illustrate the application of our models by comparing C-priority, D-priority, Hybrid batch and optimal policies for 8 months (242 days) from November 1, 2014

to June 30, 2015. Similarly, with the Dual Role Model we compare F-C-D, F-D-C, Hybrid batch and optimal policies for the same 242 days. The Hybrid batch policy is of the form $\{\text{C}_{batch2}$ if $d > 0\}$ for the Single Role Model or $\{\text{F-C}_{batch2}$ if $d > 0\}$ for the Dual Role Model. Under these Hybrid batch policies, the specialist gives priority to inpatient discharges over regular ED consultations until there are two patients waiting for ED consultation. If there are two patients waiting for ED consultation, then the specialist will complete the batch of ED consultations until the consultation queue is empty. However, the specialist will not batch ED consultations if there are no inpatient discharges.

We developed a simulation model which we use to simulate the 242 days under each policy and compare the results. The 242 days are simulated for each policy 1000 times and results are provided with the average of the 1000 replications for each policy. For the optimal policies, the action at each period is determined by solving the appropriate dynamic programming model by backward induction. At each simulated period, the backward induction algorithm is used to recursively solve equations (3.1) - (3.4) for the Single Role Model or equations (3.5) - (3.8) for the Dual Role Model.

3.6.1 Single Role Model Results

Single Role Model results are shown in figures 3-6 to 3-11. In the figures, each bar represents the average waiting and end of day penalty cost of 1000 replications for the the 242 simulated days under each policy. The base case parameters are $(w_c, w_b, w_d) = (3, 2, 1)$.

Figures 3-6 and 3-7 show Single Role Model results for the parametric analysis on $w_c = 3, 4, 5$ applied to YRH and SSRH respectively. In the base case for YRH, the D-Priority policy is closer to optimal than the C-Priority policy. As we increase the weight of w_c, C-Priority becomes closer to optimal than D-Priority. In SSRH, we see that C-Priority is always better then D-Priority. However, the magnitude of the difference changes: with higher weight to w_c, D-Priority gets worse.

Figures 3-8 and 3-9 show the results for the parametric analysis on $w_d = 1, 2, 3$ for YRH and SSRH respectively. In YRH, D-Priority is usually closer to optimal than C-Priority. This

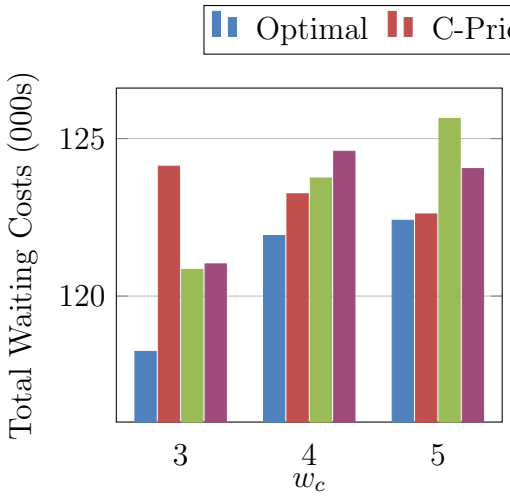

Figure 3–6: YRH, Single Role Model, $w_c = 3, 4, 5$

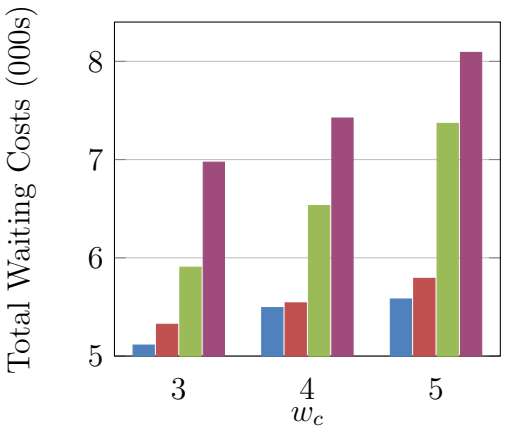

Figure 3–7: SSRH, Single Role Model, $w_c = 3, 4, 5$

seems to indicate that an early inpatient discharge strategy suggested by current guidelines is a good strategy for YRH. However, this is not always true. We also conducted a 30-day analysis for YRH for the month of April 2015. Contrary to the 8-month period, we found that the C-Priority policy achieved results that were closer to optimal than the D-Priority policy. The reason is that YRH inpatient units are commonly overloaded for extended periods throughout the winter months. In the case of an overloaded inpatient department, many consecutive days begin and stay above the boarding threshold, b^X. Consistent with our description of the structure of the optimal policy in section 4, this implies that the optimal policy D if $b > b^X$ is the same thing as the D-Priority policy in the case when an inpatient department is consistently overloaded.

In the base case for SSRH, the C-Priority policy is closer to optimal than the D-Priority policy. With increasing weight to w_d, the D-Priority policy becomes closer to optimal than the C-Priority policy only when $w_d = 3$.

Figures 3-10 and 3-11 show the results for the parametric analysis on $w_b = 1, 2, 3$. Recall that the base case has $w_b = 2$. In both YRH and SSRH, all results are sensitive to w_b. No matter what policy is adopted, higher w_b will result in higher waiting costs. As we increase w_b, we see different results at the two hospitals. In YRH, as we increase w_b, we find that

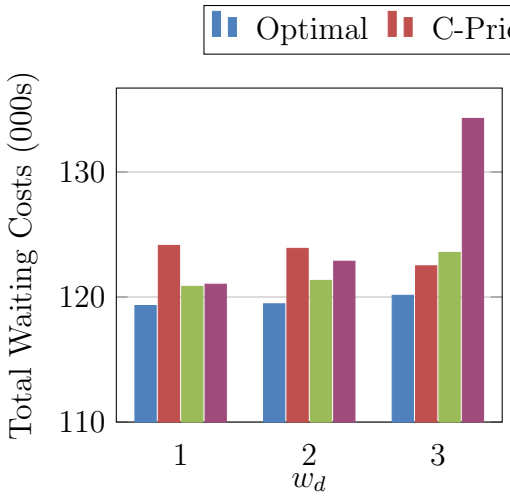

Figure 3–8: YRH, Single Role Model, $w_d = 1, 2, 3$

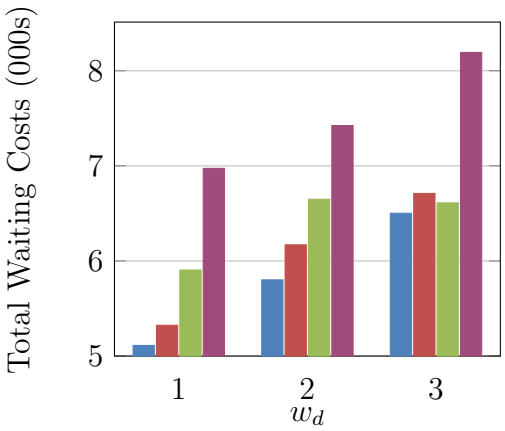

Figure 3–9: SSRH, Single Role Model, $w_d = 1, 2, 3$

the D-Priority policy is closer to optimal than the C-Priority policy. However, at SSRH, we find that the C-Priority policy is always closer to optimal than the D-Priority policy.

In figures 3-6 to 3-11, we notice that the Hybrid batch policies are sometimes closer to optimal than the C-Priority and D-Priority policies. This indicates the need to sometimes give priority to ED consultations and other times give priority to inpatient discharges. However, these hybrid batch policies are not consistently better than the C-Priority and D-Priority policies. The conditions of the optimal action depend not only on the number of patients waiting for consultation, it depends on all of the model parameters, (c, b, d).

3.6.2 Dual Role Model Results

Dual Role Model results are in Figures 3-12 to 3-17. Each bar represents the average waiting and end of day penalty cost of 1000 replications for the the 242 simulated days under each policy. The base case parameters are $(w_c, w_b, w_f, w_d) = (3, 2, 5, 1)$.

Figures 3-12 and 3-13 show Dual Role Model results for the parametric analysis on $w_c = 3, 4, 5$ applied to YRH and SSRH respectively. Lower waiting costs can be achieved with the Single Role Model rather than the Dual Role Model regardless of the specific policy adopted. As we increase the value of w_c, we observe the most profound benefit from working with a Single Role rather than a Dual Role. If a Dual Role Model was still used at YRH, the

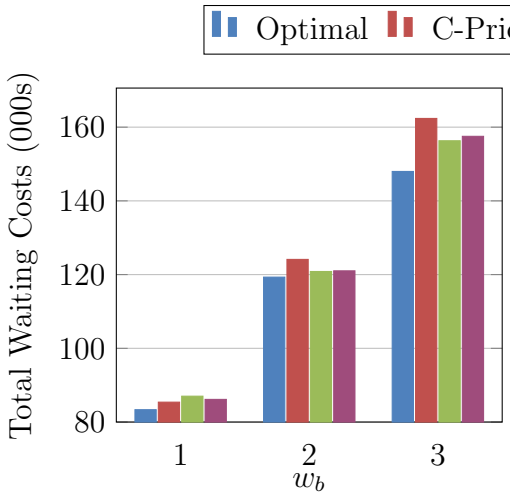

Figure 3–10: YRH, Single Role Model, $w_b = 1, 2, 3$

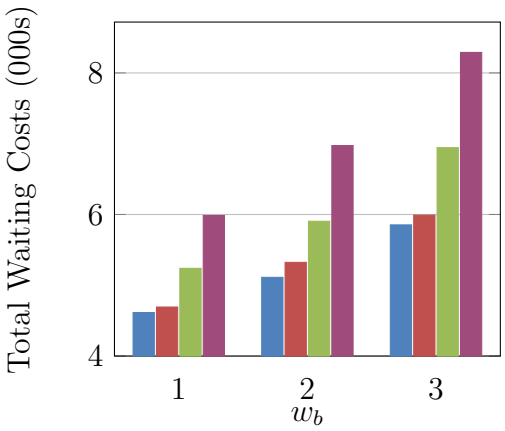

Figure 3–11: SSRH, Single Role Model, $w_b = 1, 2, 3$

F-D-C policy would achieve results closer to optimal than the F-C-D policy. In SSRH, where the Dual Role model is current practice, the F-C-D policy is always closer to optimal than the other policies. The gap between the C-Priority policy and the optimal policy becomes smaller as we increase the weight of the waiting cost for ED consultations. At SSRH, the highest waiting cost occurs if a hybrid batch policy is adopted, however batching is worse under the Single Role Model than the Dual Role Model.

Figures 3-14 and 3-15 show the results for the parametric analysis on $w_d = 1, 2, 3$ for YRH and SSRH respectively. Similar to the Single Role model for YRH with $w_d = 1$, the F-D-C policy is closer to optimal than the F-C-D policy. As we increase the weight of w_d, the F-D-C policy is much closer to optimal than the F-C-D policy with the Dual Role Model. With the Single Role Model, the results are not as sensitive to increases in w_d. Once again we note that while the results seems to indicate that an early inpatient discharge strategy suggested by current guidelines is a good strategy for YRH, this is not always true. With the 30-day analysis for the month of April 2015, we found that the F-C-D policy achieved results that were closer to optimal than the F-D-C policy at YRH. In the case of SSRH, the F-C-D policy is always better than the F-D-C policy, although the magnitude of the benefit is smaller with increasing weights to w_d.

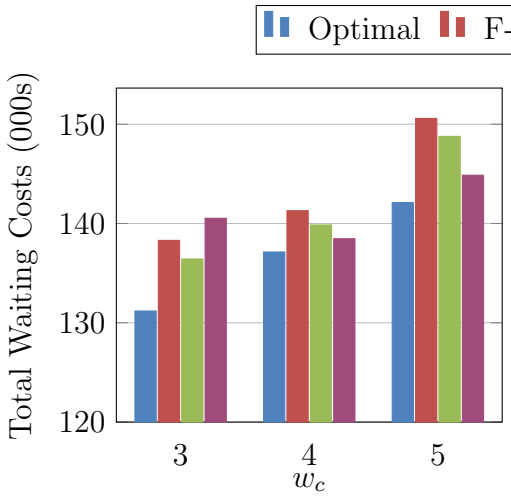

Figure 3–12: YRH, Dual Role Model, $w_c = 3, 4, 5$

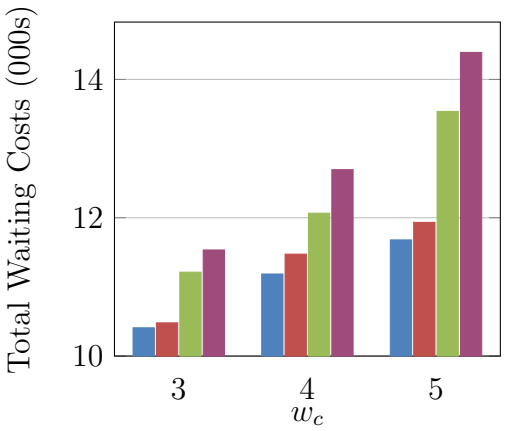

Figure 3–13: SSRH, Dual Role Model, $w_c = 3, 4, 5$

Figure 3-16 and 3-17 show results for the parametric analysis on $w_b = 1, 2, 3$ for YRH and SSRH respectively. As with the Single Role Model, in both YRH and SSRH, all results are sensitive to w_b. In YRH, the F-D-C policy is closer to optimal while at SSRH, the F-C-D policy is always closer to optimal than the F-D-C policy. We also notice that the Hybrid batch policies under the Dual Role Model are sometimes closer to optimal than the F-C-D and F-D-C policies, but not consistently better. It is also worth noting that if a hospital similar to SSRH adopts a Single Role Model, it is more important to avoid batching ED consultations than it is under the Dual Role Model.

Overall, our results confirm that YRH provides a lower waiting cost with the Single Role Model than the Dual Role Model and that SSRH could provide a lower waiting cost to patients with the Single Role Model. The results also show that an early inpatient discharge strategy suggested by current guidelines is not a good strategy for SSRH and not always a good strategy for YRH. In the case of SSRH, we observe that the D-Priority (F-D-C) policy is only better than the C-Priority (F-C-D) policy if the weight of the waiting cost for patients waiting for inpatient discharge is as high as the waiting cost for patients waiting for ED consultation. Hospitals are typically congested and concerns include health impacts for

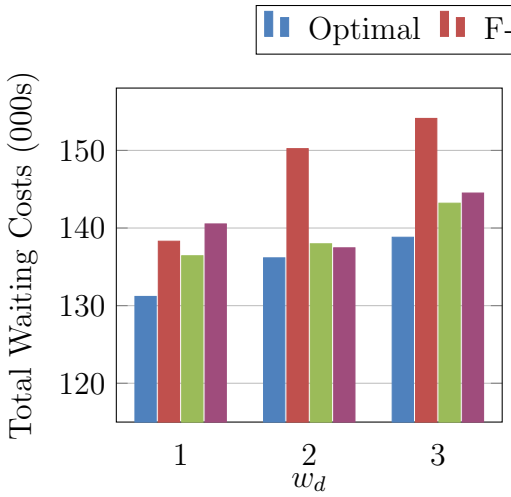

Figure 3–14: YRH, Dual Role Model, $w_d = 1, 2, 3$

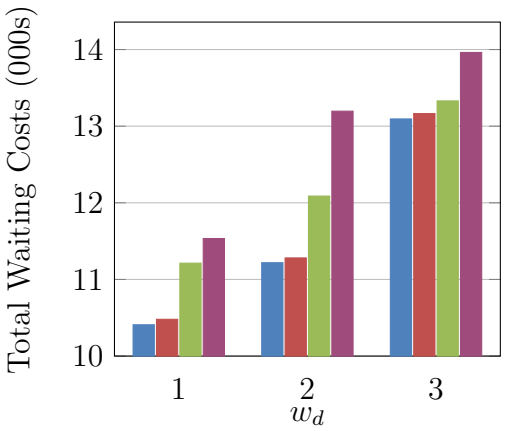

Figure 3–15: SSRH, Dual Role Model, $w_d = 1, 2, 3$

ED patients waiting for consultation. Therefore, our results with $w_d < w_c$ are more relevant in any hospital with the common challenge of ED congestion.

3.7 Discussion

With one specialist serving as the ICU physician while on call for ED consultations and inpatient ward care, prolonged waits for specialist care are inevitable. In this Dual Role situation, the Internist's morning typically begins with a queue of patients waiting in the ED for specialist consultation and a queue of patients waiting for follow-up care in the ICU. The patients in the ED arrived overnight with a condition that might require admission but the specialist was not called in to the hospital in the middle of the night for those consultations. Although the condition of those ED patients is likely not critical enough to require admission to the ICU, these patients have already been waiting in the ED for several hours for specialist consultation. Naturally, ED management's concern is that these patients will continue to occupy space and ED resources needed for new patient arrivals. On the other hand, the ICU patients are the sickest patients in the hospital and the specialist is responsible for ICU patient rounds and timely follow-up care. Ideally, specialist care would be provided to all of these patients first thing in the morning, but unfortunately this is not possible with only one specialist.

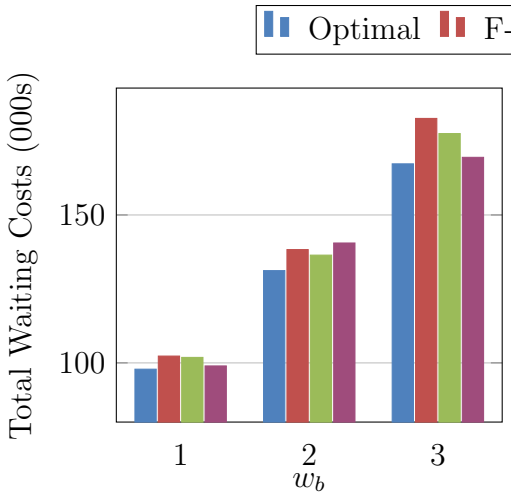

Figure 3–16: YRH, Dual Role Model, $w_b = 1, 2, 3$

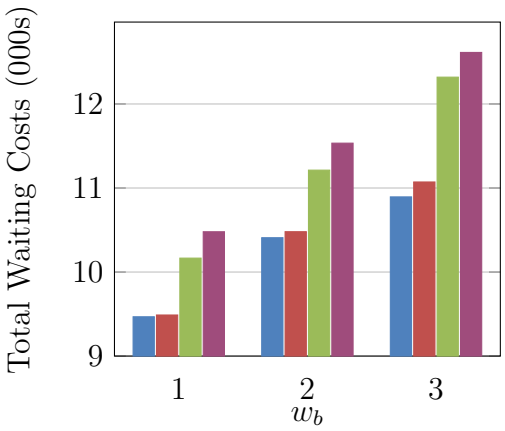

Figure 3–17: SSRH, Dual Role Model, $w_b = 1, 2, 3$

While higher patient volume contributes to a greater need for multiple Internists on call, our results show that both of our study hospitals would be better off with a separate Internist for the ICU. The greatest benefit is early in the morning, and with two to three hours of additional on call coverage at SSRH, one specialist could handle ICU rounds while another specialist clears the queue of overnight ED consultation requests. The models presented in this thesis chapter provide the ability to quantify how much better off a hospital is when specialists work with a single role rather than dual role. Furthermore, the effectiveness of different decision making policies within the single role or the dual role can be measured with these models. We believe that measurement of decisions that effect multiple hospital departments is in need of more research and we hope this study will provoke further discussion on this topic.

In practice, changing the policies adopted by physician groups can take time. In the case of YRH, it took several months of meetings among the Internal Medicine specialists before the group agreed to have a separate ICU physician. Hospital management's motivation to change from the Dual Role Model to the Single Role Model included improving Internist responsiveness for ED consultations and improving continuity of care for ICU patients. The importance of reducing fatigue among clinicians and minimizing the risk of medical errors

in intensive care is well known (Gaba and Howard 2002, Rothschild et al. 2005). With the Dual Role Model, Internists can only be on call for a maximum of three consecutive days due to the fatigue of covering both roles. In this case, ICU patient handoffs from one specialist to another are more frequent than in the Single Role Model, where the same specialist can serve as the ICU physician for a week at a time.

However, challenges for increasing on call coverage include concerns about impact on physician lifestyle, continuity of patient care and maximizing specialist utilization. For physicians interested in having a balanced lifestyle, the idea of being on call twice as often may not be an attractive proposition, and the burden is greater for smaller physician groups in rural areas. The continuity of care concern exists especially in the case of only a few hours of additional on call coverage, where one specialist would need to handoff patients to the other specialist for further care. There is evidence that adverse events occur when important information is missed during patient handoffs in Internal Medicine wards (Horwitz et al. 2008). However, if appropriate measures are taken to account for this risk, the benefits of reduced ED congestion can be achieved with minimal affect on quality of care and patient safety. There is also concern that some mornings begin without any queue of overnight ED consultations, resulting in specialist idle time that might be better spent with other patients.

Furthermore, there may be specialist compensation challenges to implementation. In some hospitals, challenges of maintaining on call coverage have resulted in more stipends being paid to some specialists for on call coverage (McConnell et al. 2007). With the Dual Role Model, Internists earn more per day than with the Single Role Model since they receive a stipend to serve as the ICU physician plus an additional on call stipend. These two stipends are paid in full regardless of the amount of work done. Internists also keep track of fee for service billing for ICU, medicine ward inpatients and ED consultations. If a daily fee for service amount exceeds the ICU stipend, then the physician gets the fee for service amount plus the on call stipend on top of that amount. This compensation mechanism is in place so that Internists do not get underpaid when they are very busy. Under the Single Role Model,

if a specialist is on call solely for ED consultations, some mornings could result in little or no work. While some specialists would welcome idle time, a different compensation strategy may be required to increase on call coverage. However, many specialists, particularly those working in rural areas, are driven less by financial incentives and more by the impact on their lifestyle or by the contribution that they can make in effectively providing health services to their communities. Nevertheless, specialist compensation could be less for some on call days and more for others, so long as it averages out in the long run.

3.8 Conclusion

In this chapter, we present two models for the workflow decisions of specialists. With the Single Role Model, we consider the situation where specialists handle ED consultations and inpatient ward care. In the Dual Role Model, we consider the more complex rural setting when the specialist has additional responsibility serving both as the ICU physician and specialist on call. The modeling efforts are particularly noteworthy in the Dual Role Model, where the process model could include a much larger state space, requiring the use of approximation methods to determine optimal actions. With the proposed models, optimal actions can be computed by backward induction. One example is the case when the specialist receives an urgent request for a high-acuity ED consultation. The decision is important but trivial: the specialist's next action will always be to address the high-acuity ED consultation request. We account for this reality in the models without the need for another state variable.

After observing every Internist working on call in two regional hospitals in rural Nova Scotia, we worked with the two physician groups to gather ED consultation and inpatient care data that is not available in hospital information systems. We also worked with decision support analysts to obtain the ED and inpatient data that is available in hospital information systems. With these data collection efforts, we were able to apply the two models, perform a parametric analysis and report on the results.

YRH is similar to an urban hospital with a separate ICU physician, so we applied the Single Role Model first. However, since YRH Internists previously worked with the typical

dual role, we also applied the Dual Role Model to YRH. We found that management's decision to staff a separate ICU physician in YRH does provide a lower waiting cost for its patients. The magnitude of this benefit depends on the relative weights of the per period waiting cost parameters for patients waiting for ED consultation, boarding for admission, and waiting for inpatient discharge. SSRH currently has the typical rural strategy with one Internist working with the dual role of ICU physician and Internist on call. Our results show that SSRH patients could benefit if the ICU physician had assistance to handle the overnight queue of ED consultations requests.

In YRH, we found that in most cases, the D-Priority (F-D-C) policy results in waiting costs that are closer to optimal than the C-Priority (F-C-D) policy. However, in SSRH, we found the reverse situation with the C-Priority (F-C-D) policy more often being closer to optimal than the D-Priority (F-D-C) policy. These results seem to indicate that an early inpatient discharge strategy suggested by current guidelines is a good strategy for YRH but not for SSRH. A closer look indicated that YRH had an overloaded inpatient department during the winter months with many consecutive days beginning and staying above the boarding threshold. In the case of a consistently overloaded inpatient department, the D-Priority policy is close to optimal. However, at other times such as the month of April 2015, an early inpatient discharge strategy is also not a good strategy at YRH.

Our analysis of the optimal policy under the Single Role Model also determines that an early inpatient discharge strategy (i.e. discharge inpatients by 11:00AM) is generally not a good strategy. Such a strategy implies that priority should be given to inpatient discharges over ED consultations which is not always true. Instead, specialists should sometimes give priority to ED consultations and other times give priority to inpatient discharges. We observe that the optimal action depends on the period number and our model state variables (c, b, d), with different forms throughout the day.

Our future research includes examining continuous time models for the workflow decisions of specialists rather than the discrete time models proposed here. We observe a

quasi-birth-and-death (QBD) structure which may be best analyzed in continuous time as an extension to the models proposed in this thesis chapter.

Chapter 4
Dialysis Facility Network Design

4.1 Introduction

Chronic kidney disease (CKD) has increasing levels of severity with the fifth and final stage classified as End Stage Renal Disease (ESRD). ESRD is treated with dialysis until transplant or death. In 2013, there were 661,648 ESRD patients in the United States. Of those ESRD patients, 29% had a functioning kidney transplant with the remaining 71% on dialysis. The 2013 annual cost for ESRD was $30.9 billion accounting for 7.1% of all U.S. Medicare costs. $27.4 billion of these costs are for dialysis patients (USRDS 2015). In Canada, the 2013 ESRD patient population was 41,913, 42% with a functioning kidney transplant with the remaining 58% on dialysis (CIHI 2015).

With hemodialysis (HD), blood is withdrawn by a machine and passed through an artificial kidney called a dialyzer. To get blood to the dialyzer, a surgeon makes an access (entrance) into the blood vessels. One tube carries blood to the dialyzer where it is cleaned and the other tube returns the cleaned blood to the patient. Conventional HD is a four hour treatment, three times per week. HD is usually done at a facility under the care of a nephrologist (kidney specialist) with nephrology nursing support. While HD uses an external machine to clean the patient's blood, peritoneal dialysis (PD) cleans blood inside the patient's body. A dialysate solution is put inside the patient's abdomen through a catheter. Waste and excess fluid is filtered into the dialysate solution which stays in the patient's belly for two hours or more. Then the cleansing fluid is drained from the patient's body through the catheter into an empty bag and discarded. PD has two methods: continuous ambulatory peritoneal dialysis (CAPD) and continuous cycling peritoneal dialysis (CCPD). With CAPD, fluid exchanges are done three to four times per day. With CCPD, fluid

exchanges are completed overnight by a machine called a cycler. PD is usually done at home after patient training is provided by nephrology nurses.

Patients travel to in-centre or satellite HD facilities three times per week or participate in home PD or home HD. Dialysis facilities tend to be clustered where there is high population density and are much harder to find in rural areas. This is true in all geographic regions of the U.S. As a result, rural patients travel much longer distances for dialysis than urban patients. Access to alternative facilities, if required, greatly increases rural patient travel time, but has little impact on urban patients. For rural patients, the average distance to the closest and second closest facility is 2.5 times and 4 times farther, respectively. The percentage of patients traveling more than 30 minutes each way to an alternative facility ranges from 2% to greater than 30% depending on the region (Stephens et al. 2013). In Canada, most dialysis patients live within 50 km of their nephrologist, but a substantial proportion (12%) live more than 150 km away.

While the likelihood of receiving a kidney transplant in Canada is not influenced by residence location (Tonelli et al. 2006), living in a remote area is a risk factor for mortality among patients receiving hemodialysis (Tonelli et al. 2007). A large international study also demonstrated that longer travel time is associated with significantly greater mortality risk and decreased quality of life for dialysis patients. The study used a sample of adult hemodialysis patients from 307 dialysis facilities in twelve countries from the Dialysis Outcomes and Practices Patterns Study (DOPPS). Patients traveling longer than 60 minutes have a 20% greater risk of death compared with those traveling 15 minutes or less. The study also found that those with the longest travel times were more likely to have problems with transportation resulting in skipped or shortened treatments (Moist et al. 2008). Dialysis is a life-saving treatment and canceling or reducing the length of treatment times clearly has a higher mortality risk. A recent study in a rural area outside New Taipei City in Taiwan showed an association between increased travel distance to dialysis units and the risk of anemia in chronic dialysis, especially among elderly patients. The study found that with

every additional km increase in travel distance, there is an increased risk of anemia (Chao et al. 2015).

Patients have the choice of dialysis mode. This choice is with the patient rather than provider since there are similar health outcomes with HD and PD (Mehrotra R et al. 2011). Regardless of the travel burden, some patients will always opt to go to an in-centre or satellite facility, while others will always opt for home dialysis. For many, the choice will vary depending on the location of available facilities. In this chapter, a case study is included for the province of Nova Scotia, where the travel burden for patients in rural areas can be greater than one hour in each direction. As a result, communities are lobbying for satellite facilities to be provided closer to home. For example, Barrington area residents presented the Nova Scotia Health and Wellness Minister with 1200 letters to request a satellite dialysis facility (Woolvett 2015a). These letters come after a patient drove into a snow bank trying to get to Yarmouth for a dialysis appointment (Woolvett 2015b).

We study the Dialysis Facility Network Design Problem (DFNDP) where the objective is to design the best possible network of dialysis facilities for all existing patients. We consider the challenge of providing reasonable patient travel time given budget and capacity management constraints and patient choice for facility-based or home dialysis. Considering that HD is a frequent treatment where patients travel three times per week, shorter travel times are ideal from a patient welfare perspective. However, since the average cost for home dialysis is less than the average cost for facility-based dialysis, healthcare management may prefer home dialysis from a cost perspective. If a satellite facility is created, will it reduce participation in home therapy? While establishing satellite facilities requires an initial setup cost due to construction, home dialysis also require a setup cost for home renovations and also for patient training before dialysis can be done at home. We develop the DFNDP model to identify the optimal location of dialysis facilities and determine the capacity requirements for dialysis stations at the facilities. We provide a general model that can be used for the establishment of a new network of dialysis facilities. Alternatively, with a few additions

the model can be used to consider a subset of the DFNDP. For example, in the case study we focus on the problem of identifying the optimal network of satellite facilities without changing the locations of an existing network of in-centre facilities. In that case, we consider the possibility to establish new satellite facilities and/or expand capacity at the in-centre facilities beyong the minimum capacity required for acute care purposes.

In the remainder of the chapter, we begin with a brief history of dialysis treatment in section 4.2, review the most relevant literature in section 4.3, and define our mathematical model for the DFNDP in section 4.4. We then describe our study setting and data sources in section 4.5, report results of the application of the DFNDP in a case study in section 4.6, and summarize our findings with concluding remarks in section 4.7.

4.2 A Brief History of Dialysis Treatment

Dialysis as a treatment for *acute* renal failure first became possible in 1943 with the development of the artificial kidney by Willem Kolff in Kampen, Netherlands (Kolff et al. 1944). From March 17, 1943 to July 27, 1944, fifteen patients were treated with the artificial kidney and one survived. For the patient that survived, the patient's wife believed that the artificial kidney saved his life, but Dr. Kolff did not think so. The development and experiments with the artificial kidney occurred during World War II (WWII). To protect the risk of losing all eight artificial kidneys that he made from bombing, Kolff hid the artificial kidneys in different parts of Kampen. After WWII, Kolff sent one artificial kidney to the Hammersmith Hospital in London, one to Mount Sinai Hospital in New York, and one to the Royal Victoria Hospital in Montreal (Kolff 1965).

Long-term dialysis for *chronic* renal failure first became possible with the development of the Teflon shunt by Belding Scribner and colleagues at the University of Washington in 1960 (Quinton et al. 1960). The Seattle Artificial Kidney Center was the first outpatient dialysis center which opened in 1962. Many of the important medical developments for dialysis followed in Seattle in the 1960s. The first chronic dialysis patient in Seattle was Clyde Shields, a Boeing machinist, who died from a heart attack after eleven years on

dialysis. Dialysis also worked for subsequent patients including the first three chronic dialyis patients who all lived for more than ten years on dialysis. However, demand for the new life-saving therapy far exceeded supply, and the first bioethics committee was established to determine who shall live. The patient selection committee included one physician (not a nephrologist), one lawyer, one housewife, one businessman, one labor leader, and one minister. The committee was in place for a decade until 1972 when all patients referred for dialysis treatment could be accepted, making the committee no longer necessary. Note however, that at that time, referrals for dialysis included only 10% of patients aged 56 years or older and few diabetics were accepted then.

Home HD was initiated in Boston in 1963, then in London and Seattle in 1964. When the bioethics selection committee turned down a 15 year old girl with renal failure, Dr. Scribner worked with Professor Babb and colleagues to make a single patient automated machine which made home HD feasible. Caroline, whose father was a friend of Professor Babb, became the first Seattle home HD patient who lived on home dialysis for six years. This single patient automated machine developed for home HD turned out to be the prototype for single patient equipment which is still used today for facility-based HD.

Other important medical developments include the replacement of the Scribner shunt with a surgically created arteriovenous fistula (AVF) in 1966 (Brescia MJ et al. 1966). Success of chronic dialysis depends on repeated access to blood vessels to provide continuous flow of up to 250 to 300 ml. per minute. While the external Teflon shunt did make chronic dialysis possible, it presented clinical and psychological problems which were eliminated with AVF. Medical and technical developments in PD also occurred in the 1960s and 1970s. The first automated PD equipment and peritoneal access devices were developed when Dutch PD pioneer Fred Boen came to Seattle in 1962. The peritoneal catheter was then developed by Henry Tenckhoff in 1968 (Tenckhoff and Schechter 1968) after he came to Seattle and continuted to work on PD when Dr. Boen returned to the Netherlands in 1963. Another important development in PD occurred with the introduction of continuous ambulatory

peritoneal dialysis (CAPD) in 1978 (Popovich et al. 1978). For further details on the history of dialysis, see Blagg (1999) and Blagg (2007).

4.3 Literature Review

The operations management (OM) literature includes models of transplant wait lists and organ allocation for kidney (Zenios et al. 2000) and liver (Alagoz et al. 2007). OM studies on therapy initiation and management include liver transplant timing (Alagoz et al. 2004), HIV (Shechter et al. 2008), and dialysis therapy (Lee et al. 2008). Recently, Skandari et al. (2015) developed a continuous time dynamic programming model for vascular access choice for HD. AVF has lower infection and mortality rates than a central venous catheter (CVC), however AVF surgery needs to be done three months in advance and it may be unsuccessful. On the other hand, CVC can be used immediately after a patient begins HD. The authors found that the optimal policy to maximize quality-adjusted life expectancy is immediate AVF surgery up to a threshold time, after which CVC becomes the optimal vascular access choice.

The health economics and health services literature on patient preferences is also noteworthy. Early studies include identifying travel time as a barrier to access hospital care (Bosanac et al. 1976) and examining health care priorities among rural patients (Kane 1969). More recent patient choice studies use multinomial logit (MNL) models for hospital choice among rural patients (Adams et al. 1991, Tai et al. 2004, Roh et al. 2008) and cataract patients (Sivey 2012). Another example is a choice-based survey used to identify attributes affecting hospital choice among patients from five Canadian teaching hospitals (Cunningham et al. 2008). However, these studies do not consider patient preferences within the context of a facility network design problem.

Relevant studies on facility network design include models for preventive care (Verter and Lapierre 2002, Zhang et al. 2009, 2010, 2011), primary care (Graber-Naidich et al. 2014, Parker and Srinivasan 1976) and access to public sector services (Aboolian et al. 2015). From a methodological perspective, our model includes some elements that are common with the

preventive care models proposed in Zhang et al. (2011). In that study, the objective is to design a network of preventive care facilities to maximize participation in a preventive care program. The authors apply their models to breast cancer screening for the city of Montreal and analyze the impact of client choice on the facility network. In the preventive care setting, clients have the choice of which facility to patronize. In that study, the authors assume that proximity to facility is the main attractiveness feature and apply an MNL model for the patient choice function. In the case of dialysis, the objective function is different since all patients who wish to survive will participate in some mode of dialysis. However, since patients can choose dialysis mode (i.e. to go to a facility or home dialysis), we have an interesting patient choice problem included in the DFNDP.

To our knowledge, there is only one previous study on dialysis facility network design. In their early work, Eben-Chaime and Pliskin (1992) considered two models: 1) minimize costs subject to welfare constraints and 2) maximize patient welfare subject to resource constraints. Consideration for patient welfare is consistent with the model proposed in this chapter, however, our model also considers patient choice for dialysis mode. Home dialysis and satellite facilities are not considered by Eben-Chaime and Pliskin (1992) and their objective function minimizes a simple travel time measure such as the longest travel time. Furthermore, the model proposed in this chapter has a novel objective function that considers the travel time for all patients with greater weight given to minimizing longer travel times. Our model is designed so that the best possible solution can be determined for the dialysis facility network given available resources.

4.4 Dialysis Facility Network Design Problem (DFNDP) Model

We develop a mathematical model for the dialysis facility network design problem (DFNDP), considering the impact of patient choice for dialysis mode. We incorporate the challenges of capacity management and budget constraints required to find a feasible solution. We identify the optimal network of facilities using a patient welfare objective to provide reasonable travel time to all patients.

Consider the problem of planning a network of dialysis facilities for patients residing in n health regions. I^r represents the set of patient residential locations in health region r and there are J potential facility locations. We define two types of decision variables: 1) location decisions, y_{jk}, if dialysis facility is opened at location j with k dialysis stations and 2) allocation decisions, x_{ij}, if patient i is allocated to facility j. However, with the DFNDP, patients can also choose home dialysis, and we use index $J+1$ if the allocation decision is for home dialysis. With respect to patient choice, we have $I_1^r \subset I^r$ for patients who prefer to go to a facility regardless of travel time, $I_2^r \subset I^r$ for patients who prefer home dialysis regardless of travel time, and $I_3^r \subset I^r$ for patients whose choice for dialysis mode depends on travel time. For patients in I_3^r, we assume that patients will choose home dialysis if their travel time exceeds a threshold time U.

Budget and cost are a part of the DFNDP as well, and we set our cost parameters according to a multi-year planning horizon, e.g. 30 years. In this general model, we assume that there are no existing facilities, so the budget parameter, B, represents the total amount available to build and operate facilities and provide dialysis to patients over the planning horizon. If there are existing facilities, the DFNDP can be extended as we have done in the case study, for example, to plan for a set of new satellite dialysis facilities, given an existing set of in-centre facilities. In that case, the cost to build new facilities, operate existing and new facilities and provide dialysis to patients over the planning horizon can be considered. We also assume that patient locations are a steady state over the planning horizon, such that if a patient dies or receives a transplant, a new patient at (almost) the same residential location develops ESRD and requires dialysis such that the costs would be continuous over the planning horizon.

DFNDP model parameters are given in Table 4-1.

Table 4–1: DFNDP Model Parameters

$r = 1, ..., n$	Index for n health regions
I^r	Set of patient residence locations in health region r
$i \in I^r$	Index for patient residence locations in health region r
$j = 1, .., J, J + 1$	Index for J potential facility locations, including $J + 1$ for home dialysis
$k = 1, ..., K$	Index for K potential dialysis stations at a facility
$I_1^r \subset I^r$	Set of patient residence locations in health region r for patients who prefer to go to a facility regardless of travel time
$I_2^r \subset I^r$	Set of patient residence locations in health region r for patients who prefer home dialysis regardless of travel time
$I_3^r \subset I^r$	Set of patient residence locations in health region r for patients whose choice for dialysis mode is unknown
h_r	Number of patients residing within health region r
t_{ij}	Travel time from patient residence i to facility j
T	Target for reasonable travel time
U	Threshold travel time for patients whose choice for dialysis mode is unknown
w	Number of patient treatment time slots per dialysis station per week
B	Budget
F_{jk}	Cost per k-station facility at location j
V_j	Cost per patient at location j
x_{ij}	1 if patient i is allocated to facility j, 0 otherwise,
y_{jk}	1 if facility j has k dialysis stations, 0 otherwise.

$$min \sum_{j=1}^{J} \sum_{i=1}^{I} x_{ij}(e^{t_{ij}-T}) \tag{4.1}$$

s.t.

$$\sum_{j=1}^{J+1} x_{ij} = 1, \qquad i = 1, ..., I, \tag{4.2}$$

$$\sum_{k=1}^{K} y_{jk} <= 1, \qquad j = 1, ..., J, \tag{4.3}$$

$$x_{ij} <= \sum_{k=1}^{K} y_{jk}, \qquad i = 1, ..., I, j = 1, ..., J, \tag{4.4}$$

$$t_{ij}x_{ij} <= t_{ip} + G(1 - \sum_{k=1}^{K} y_{pk}), \qquad i = 1, ..., I, j, p = 1, ..., J, \tag{4.5}$$

$$\sum_{j=1}^{J+1} \sum_{i=1}^{I^r} x_{ij} = h_r, \qquad r = 1, ..., n, \tag{4.6}$$

$$x_{i,J+1} = 0, \qquad i \in I_1^r, r = 1, ..., n, \tag{4.7}$$

$$x_{i,J+1} = 1, \qquad i \in I_2^r, r = 1, ..., n, \tag{4.8}$$

$$t_{ij}x_{ij} <= U, \qquad i \in I_3^r, j = 1, ..., J, \quad r = 1, ..., n, \tag{4.9}$$

$$\sum_{i=1}^{I} x_{ij} \leq w \sum_{k=1}^{K} k y_{jk}, \qquad j = 1, ..., J, \tag{4.10}$$

$$\sum_{j=1}^{J} \sum_{k=1}^{K} F_{jk} y_{jk} + \sum_{i=1}^{I} \sum_{j=1}^{J+1} V_j x_{ij} \leq B, \tag{4.11}$$

$$x_{ij} = \{0, 1\}, \qquad i = 1, ..., I, j = 1, 2, ..., J,$$

$$y_{jk} = \{0, 1\}, \qquad j = 1, ..., J, k = 1, 2, ..., K.$$

Our objective in (4.1) is to minimize long travel times to improve patient welfare. Given a target T for reasonable travel time, our objective is to minimize patient travel times above the target. If the travel time, $t_{ij} = T$, then $e^{t_{ij}-T} = e^0 = 1$. If the travel time $t_{ij} < T$, then $e^{t_{ij}-T}$ will be a small fractional value between zero and one. On the other hand, if the travel time $t_{ij} > T$, then $e^{t_{ij}-T}$ will be a large value with larger values of t_{ij} resulting in exponentially larger values. These large values represent the very long travel times which we seek to avoid to improve patient welfare. The longer the travel time the more we seek to avoid them. Constraints (4.2) ensure that all dialysis patients will be allocated to one

of the J facilities or participate in home dialysis ($J + 1$). Constraints (4.3) ensures that only one capacity configuration can be located in facility j. For each facility, at most one of the y_{jk} values can be one, so we have at most one facility located at location j. The number of dialysis stations at facility j is represented by k, so if $y_{jk} = 1$, the location and capacity decision is to open a k-station facility at location j. If $\sum_{k=1}^{K} y_{jk} = 0$ for facility j, then there will be no facility opened at location j. Constraints (4.4) ensure that patients are allocated only to open facilities. Constraints (4.5), where G denotes a big number, indicate that patients who are allocated to a facility choose the closest open facility. G can be set to the longest travel time in the network.

Constraints (4.6) indicate that all patients in each health region are allocated to a facility or home dialysis. Constraints (4.7) ensure that allocation decisions for the subset of patients from health region r who choose to go to a facility regardless of travel time are not allocated to home dialysis. Similarly, constraints (4.8) ensure that allocation decisions for the subset of patients from health region r who choose home dialysis regardless of travel time are allocated to home dialysis. For the remaining patients in health region r, constraints (4.9) represent the patient choice function for dialysis mode for the remaining patients whose choice depends on travel time. For the remaining I_3^r subset of patients for health region r, the corresponding allocation decisions are determined by constraints (4.9) together with constraints (4.2), (4.4) and (4.5). For example, if a patient's travel time is greater than U, then the allocation decision will be for home dialysis according to constraints (4.2) and (4.9). Otherwise, the patient will be allocated to the closest open facility according to constraints (4.2), (4.4), (4.5) and (4.9).

Constraints (4.10) are for capacity management and constraint (4.11) is the budget constraint. Constraints (4.10) ensures that for each facility j, the supply of dialysis machines is sufficient to handle the current demand for dialysis. With facility-based HD, each patient requires 4 hour treatment, 3 times per week occupying a dialysis machine 12 hours per week. For example, in Yarmouth Regional Hospital, HD patients have treatment time

slots on Monday-Wednesday-Friday mornings (M-W-F AM), Monday-Wednesday-Friday afternoons (M-W-F PM), Tuesday-Thursday-Saturday Mornings (Tu-Th-Sa AM), or Tuesday-Thursday-Saturday afternoons (Tu-Th-Sa PM). If the dialysis machine operating time is 8 hours per day, 6 days per week totalling 48 hours per week, then each dialysis machine can be used for $w = 4$ patient treatment time slots per week. The budget constraint (4.11) includes F_{jk}, the cost to open and operate a k-station facility at location j, and costs per patient included in V_j. For home dialysis, since there is only one facility per patient, we incorporate all setup and ongoing operating costs in V_{J+1}. However, in the case of in-centre or satellite facilities, costs that are incurred independent of the number of patients receiving treatment are included in F_{jk}, while per patient treatment costs are separated and included in V_j.

The proposed model is a mixed integer program (MIP) that can be solved with a standard commercial MIP solver such as CPLEX. In the next two sections, we demonstrate the application of the proposed DFNDP model.

4.5 Study Setting and Data Sources

Our study setting is the Canadian province of Nova Scotia. The largest population centre is the Halifax Regional Municipality (HRM), an urban centre with almost half of the province's residents. Halifax is physically located in the centre of the province. Nova Scotia also has many smaller rural communities, comprised mostly with an aging population with growing needs for healthcare services. There are 39 hospitals located throughout the province, however, most do not have dialysis facilities. Nova Scotia has had five in-centre renal dialysis facilities for adults and nine satellite dialysis facilities. One in-centre facility is located in the Western Zone at Yarmouth Regional Hospital, two are in the Central Zone within HRM at the QEII Health Sciences Centre in Halifax and Dartmouth General Hospital, and two are in the Eastern Zone at Cape Breton Regional Hospital in Sydney and Northside General Hospital in North Sydney. Figure 4-1 provides a map with the locations of in-centre facilities numbered 1 to 5 and satellite dialysis facilities numbered 6 to 14.

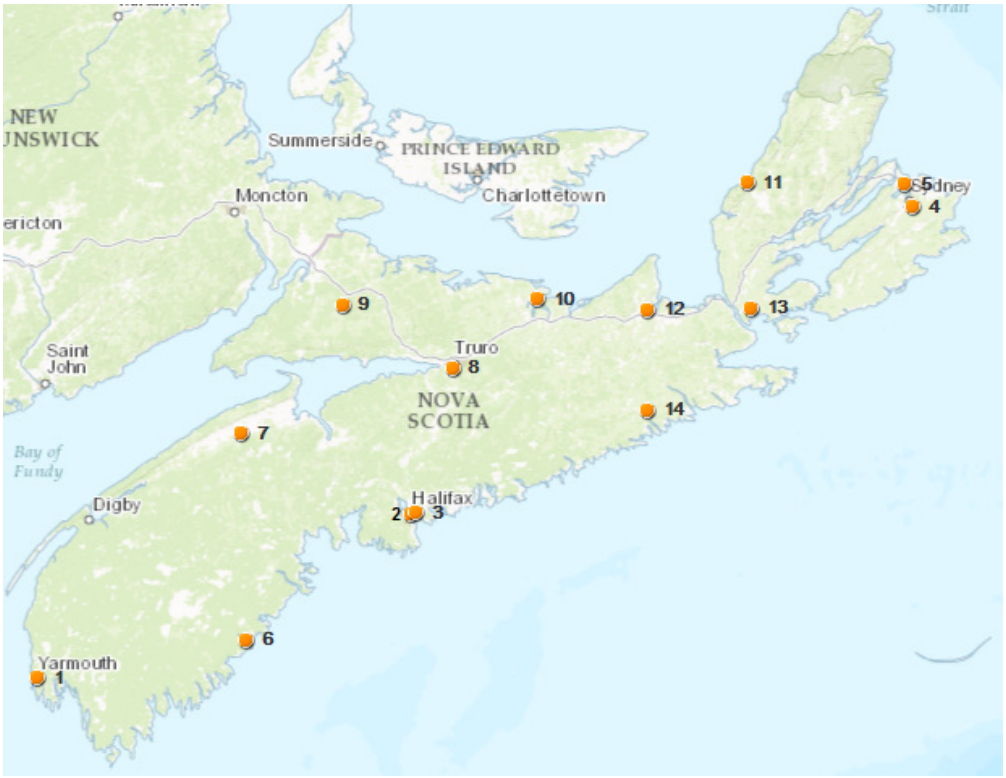

Figure 4–1: Nova Scotia Dialysis Facility Locations

The Nova Scotia Renal Program (NSRP), established as a provincial program of the Nova Scotia Department of Health and Wellness, has the mandate to improve renal health and care for *all* Nova Scotians. In rural areas, the travel burden to dialysis facilities can be greater than one hour in each direction. Communities are lobbying for facilities closer to home to relieve the travel burden for patients who require hemodialysis three times per week for each four hour treatment. The NSRP currently reviews requests from rural communities to determine whether or not to add a new satellite dialysis facility. Factors influencing these decisions include patient travel times as well as the level of adoption of home dialysis. When home dialysis adoption is high, there may be a reluctance to create a satellite dialysis facility closer to patient homes due to the possibility that fewer patients will choose home dialysis.

Ethics approval was obtained from the Nova Scotia Health Authority (NSHA) Research Ethics Board (REB), file no.: 1020103 and the McGill REB, file no.: 415-0316. Our data requirements include 1) patient preferences for dialysis mode 2) patient locations and potential facility locations and 3) budget and cost data.

4.5.1 Patient Preferences

We obtain patient preferences through a Dialysis Patient Survey, shown in the Appendix. We include all dialysis patients from the Western Zone's Renal Program as potential participants for the patient survey. There were 59 patients on dialysis within the Western Zone during the study timeframe. We excluded two patients with dementia and two patients signed a consent form but died before completing the survey. Out of the remaining 55 possible participants, we obtained 47 completed surveys resulting in an 85.5% participation rate. 63% of the survey participants are male. The mean age of all participants is 67.4 years old with standard deviation of 10.9 years. For the survey participants, the mean travel time to HD facility is 39.5 minutes with standard deviation of 23.9 minutes.

Training for home dialysis is currently only offered in Halifax. The survey results showed that patients from the Western Zone would not switch to home dialysis if training remains only offered in Halifax. However, if training for home dialysis is offered in Yarmouth, the survey results indicate that some patients would switch to home dialysis. Considering the responses that assume that training for home dialysis is offered in all in-centre HD facilities including Yarmouth, 42.6% of patients always prefer HD (in-centre or satellite) regardless of travel time, 12.8% always prefer home dialysis, and the choice depends on travel time for the remaining 44.6% of participants.

Survey responses also revealed some of the reasons that patients do not use home dialysis, including:

> Do not have room at home.
>
> I have iron in my [well] water.
>
> I prefer going to a centre. Home dialysis would cause too many anxieties for me.

I don't want to because it is too much work for me and to do a full time job.

[The patient] is living [alone]. I think [the patient] has a hard time understanding instruction.

I do not want to [go to] Halifax for training.

I do not want the responsibility of operating a dialysis machine.

Traveling to hospitals for issues such as peritonitis would create a hardship for me and my family. One hospital is 25 miles away, the other with dialysis support is about 60 miles away.

Our water is provided by a well, our drainage is a personal septic tank.

My [spouse] is elderly and would add stress.

4.5.2 Patient Locations and Potential Facility Locations

We requested province-wide data from the Canadian Institute for Health Information (CIHI) in July 2015. CIHI maintains ESRD data within the Canadian Organ Replacement Registry (CORR). At the begining of 2016, CIHI determined that the scope of the request could be covered under the Graduate Student Data Access Program (GSDAP). We obtained signatures from required McGill personnel including the Chief Technology Officer (CTO). A laptop was prepared with full disk encryption and a secure case was purchased for the laptop to satisfy CIHI's security requirements. In August 2016, CIHI and the privacy branch of the Nova Scotia Department of Health and Wellness decided not to authorize the release of data for this study.

Subsequent efforts included developing the following methodology for generating location data for province-wide patient residences. First, we begin with the number of dialysis patients for each health region. The number of dialysis patients by district health authority (DHA) in fiscal year 2011-2012 is publicly available on the NSRP website. This information is shown in Table 4-2. Next, we obtain population density from the Canadian census which is publicly available on the Statistics Canada web site. The 2011 census includes population data for more than 15000 dissemination blocks in Nova Scotia, each geocoded with X,Y coordinates. After associating the corresponding DHA for each dissemination block,

we can assign appropriate weights for each dissemination block based on census population with the resulting weights summing to 1. For each patient, we used the *rand*() function in Excel (Microsoft, Redmond, WA) to generate a random number between 0 and 1. The dissemination block that is closest to the random number is assigned as the patient residence location. Repeating for all of the 581 dialysis patients that were in Nova Scotia in fiscal year 2011-2012, we generated patient residence locations as depicted in figure 4-2.

Table 4–2: Number of Dialysis Patients in Nova Scotia by District Health Authority (DHA)

District Health Authority (DHA)	In-Centre and Satellite Facility	Home Dialysis
South Shore	30	8
South West	40	7
Annapolis Valley	28	8
Colchester East Hants	25	8
Cumberland	13	< 5
Pictou County	20	< 5
Gusborough Antigonish Strait	16	5
Cape Breton	97	14
Capital	220	37
Total	489	92

For potential facility locations, we use 40 locations: all of the 39 existing hospital locations in Nova Scotia and one existing satellite dialysis facility location. New satellite facilities could be created elsewhere too. Therefore, we tested the model with additional potential facilities initially. However, we found that the base case of the optimal solution selected a subset of the 40 locations so we focused on these 40 potential facilities for this study. These 40 locations are shown in figure 4-3.

4.5.3 Budget and Cost

We worked with NSHA Renal Program Management for budget and cost information. In the medical literature, dialysis cost analysis studies typically compute an overall average annual cost per patient by modality. Examples of dialysis cost analysis studies include Lee et al. (2002), Mcfarlane et al. (2002), Klarenbach and Manns (2009), and Komenda et al. (2010). Similar to these studies, NSHA Renal Program Management provided average annual cost per patient by modality as shown in Figure 4-4. Note that these costs include

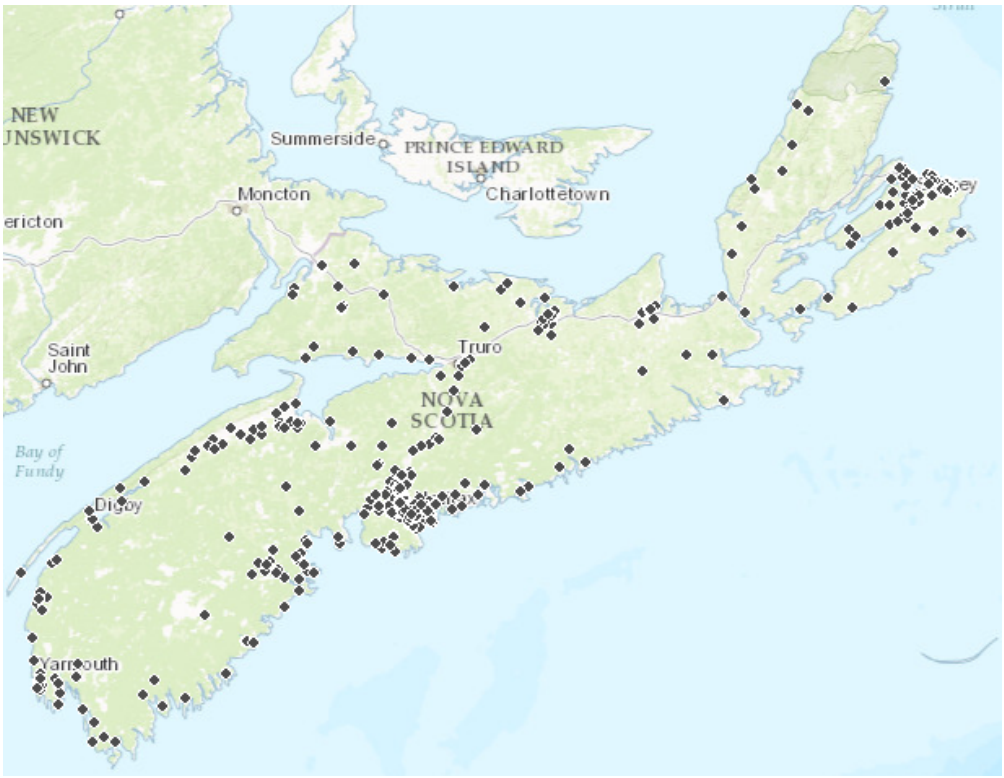

Figure 4–2: Generated Patient Locations

both per treatment costs, including dialysis supplies, but also include costs related to operating the facility. However, construction costs to build a new satellite facility are excluded from Figure 4-4. The largest portion of the cost comes from staffing and dialysis supplies. Consistent with the cost analysis studies in the medical literature, home dialysis appears to be considerably cheaper overall and in-centre HD appears to be the most expensive modality overall. Although the cost for dialysis supplies is much higher for home dialysis, this additional cost is offset by a greater reduction in other costs.

To incorporate satellite facility construction costs, we consider the cost over a 30 year horizon. Construction costs are incurred initially but the satellite facility can be used for many years in the future. An example of the total cost for a 6-station facility or home dialysis over a 30 year horizon is shown in Figure 4-5. These costs include both the costs to operate the facility regardless of the number of patients as well as the cost per patient. The per patient cost is calculated for 24 patients over the 30 year horizon, assuming that

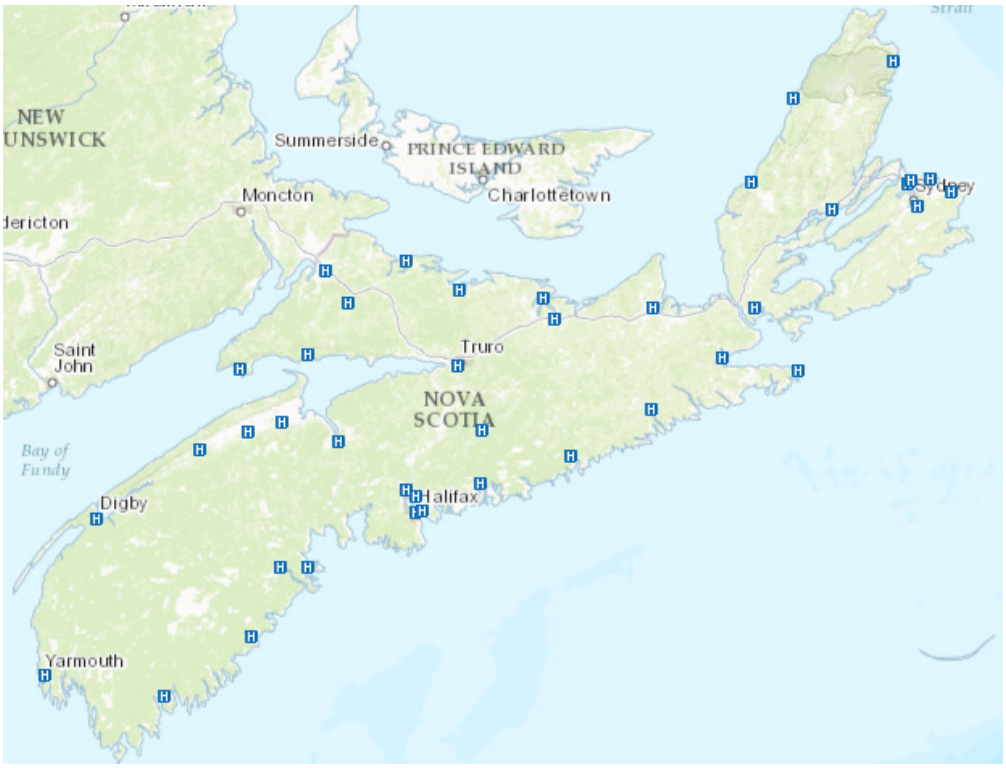

Figure 4–3: Potential Facility Locations

the 6-station facility has four shifts per week and there are always patients using the dialysis stations during the four shifts. In-Centre HD has the highest cost, followed by Satellite HD and Home Dialysis respectively.

However, in order to use the cost data in our model, we need to consider the cost per facility and cost per patient *separately*. The facility location decision is to have a dialysis facility or not at a potential location and the costs associated with this decision are incurred regardless of the number of patients that use the facility. The treatment cost per patient relate to the patient allocation decisions to open facilities and capacity management plays a role. For example, if a 6-station facility has 20 patients allocated to that facility, the facility cost is incurred and the per patient cost is incurred for 20 patients. If the 6-station facility is at full capacity, then the facility cost is the same but the per patient cost is for those 24 patients. We separate the cost per facility and per patient cost accordingly, providing an example of the 6-station facility cost in figure 4-6 and the per patient cost in figure 4-7. For the

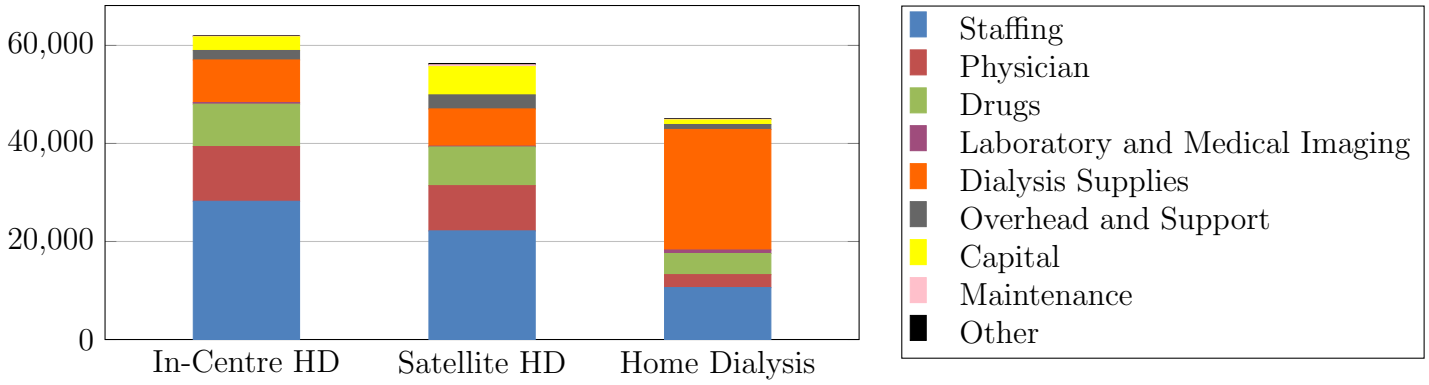

Figure 4–4: Average Annual Cost per Patient (including facility cost, except construction)

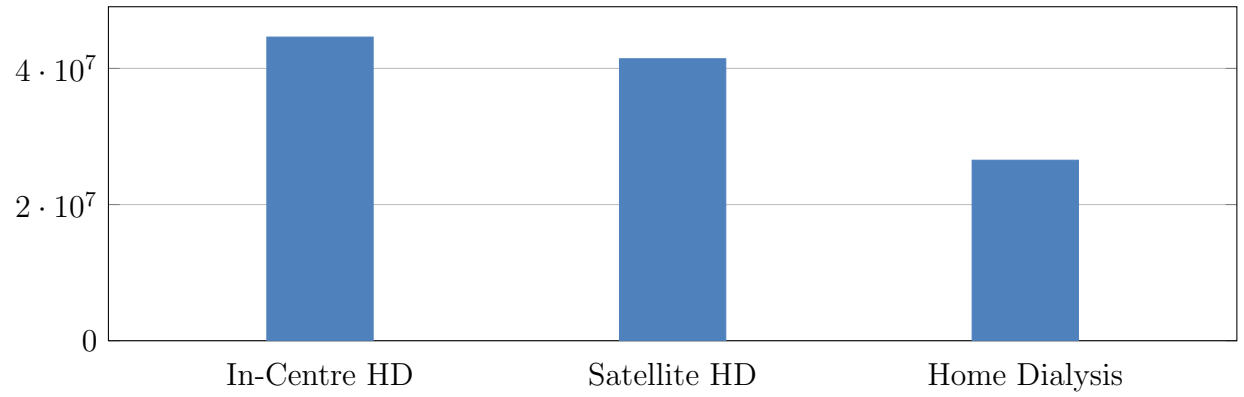

Figure 4–5: Total cost for 24 patients at 6-station facility or home (30 year horizon)

facility costs, we assume that staffing (excluding physician), construction, capital (including dialysis stations), overhead, support, and maintenance costs are incurred regardless of the number of patients allocated to the facility. We observe that staffing costs are the largest component of the facility costs with in-centre facilities having considerably higher staffing costs. In-centre facilities are already built so construction cost only applies to satellite facilities. Considering that home dialysis supports only one patient, we include all costs for home dialysis in the per patient cost in figure 4-7. In the case of per patient costs depicted in figure 4-7, we have dialysis supplies, physician treatment, drugs, laboratory and medical imaging and other for all dialysis modes. For home dialysis we also include operation costs including 1) the cost of home renovations required to support dialysis at home, 2) staffing costs due to training patients to use home dialysis, 3) capital (including dialysis stations),

and 4) overhead, support, and maintenance. The home renovations can include electrical and plumbing requirements to handle the additional water needed to support dialysis. These costs are reimbursed by the healthcare provider. We assume that the home renovation and staffing costs for home dialysis training are incurred six times over the 30 year horizon. We assume that the cost is incurred six times over a 30 year horizon since five-year patient survival rates are commonly reported for dialysis patients. While home dialysis is usually more expensive overall, the per treatment cost is more than the other modalities mainly due to higher dialysis supply costs. As a result, if an existing facility has available capacity, it may be cheaper to allocate a patient to a facility since the facility costs have already been incurred. Therefore, home dialysis is not always the cheapest modality.

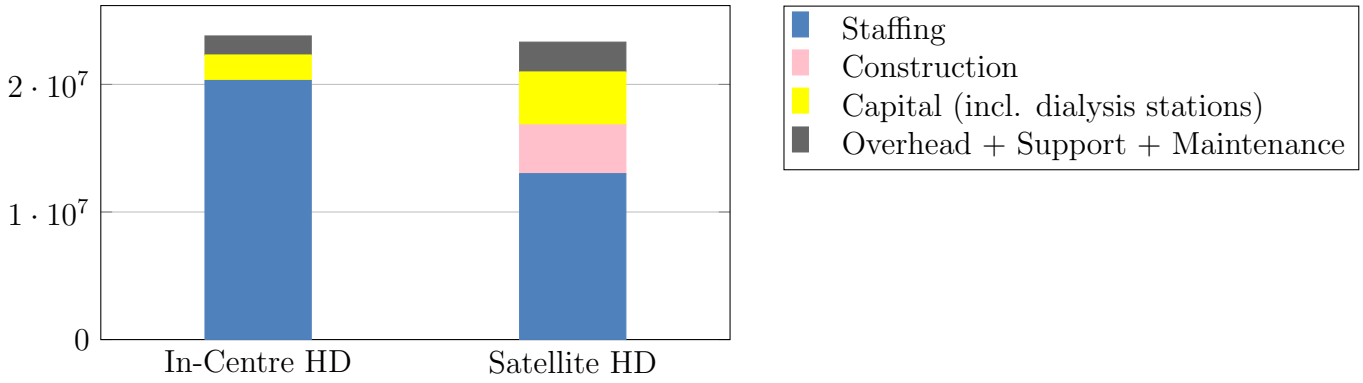

Figure 4–6: 6-station Facility Cost (30 year horizon)

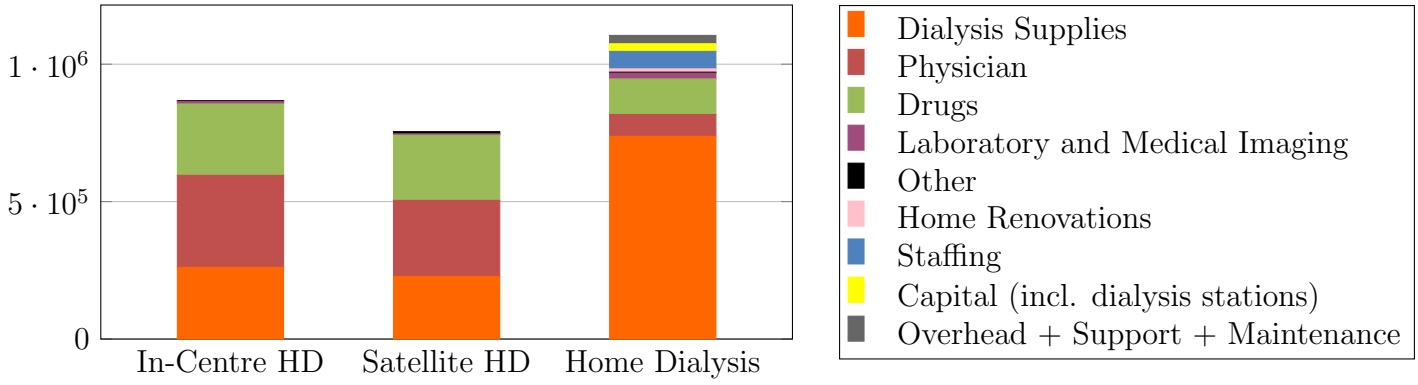

Figure 4–7: Cost per Patient excluding Facility Cost (30 year horizon)

The budget for our case study is the sum of all facility and patient costs that would be incurred over a 30 year horizon for the existing dialysis facilities. We obtained the number of stations at each existing facility so that capacity costs are also incorporated.

4.6 Case Study

In this case study, we consider the DFNDP applied to Nova Scotia assuming that there are no existing satellite facilities. We do this in order to gain a better understanding on the optimal location and capacity decisions for satellite facilities. However, considering that in-centre facilities have additional acute care requirements, we assume that all existing in-centre facilities are open at the current locations. We did not have information on the proportion of in-centre facilities that need to be reserved for acute care purposes, so we assume that the existing in-centre capacity is the minimum required capacity. Defining S_j as the minimum required capacity for each in-centre facility, $j = 1, ..., 5$, we can add the following constraints to the DFNDP for our case study:

$\sum_{k=1}^{K}(ky_{jk}) \geq S_j, j = 1, ..., 5$

The list of DFNDP tests is provided in Table 4-3. Our base case has 1) the target for reasonable travel time, $T = 45$ minutes, 2) the threshold for patients whose choice depends on travel time, $U = 45$ minutes and 3) the budget, B =base budget. We then run a parametric analysis on $T = 30, 45, 60, 75, 90$, $U = 30, 45, 60, 75, 90$ and B =base budget + 2%, + 4%, + 6%, + 8%, and + 10% respectively.

In order to quantify and compare our DFNDP solutions to our test case and the existing dialysis facility network, we provide the objective value, home dialysis proportion, maximum, mean, and standard deviation for patient travel time, t_{ij}. We also calculated the number of patients with $t_{ij} > T$. These results are provided in Table 4-4. To obtain the "Existing Facilities" results, we use our generated patient locations and manually allocate patients to the existing facilities (1 to 14). We used the same parameters as the base case for the DFNDP model, $T = 45, U = 45$. Consistent with the DFNDP model, we attempted to

Table 4–3: DFNDP Tests

Name	T	U	B
base case	45	45	base budget
$T = 30$	**30**	45	base budget
$T = 60$	**60**	45	base budget
$T = 75$	**75**	45	base budget
$T = 90$	**90**	45	base budget
$U = 30$	45	**30**	base budget
$U = 60$	45	**60**	base budget
$U = 75$	45	**75**	base budget
$U = 90$	45	**90**	base budget
budget + 2%	45	45	**base budget + 2%**
budget + 4%	45	45	**base budget + 4%**
budget + 6%	45	45	**base budget + 6%**
budget + 8%	45	45	**base budget + 8%**
budget + 10%	45	45	**base budget + 10%**

allocate patients to the closest facility, but this was not always possible as there is insufficient capacity to meet that constraint. We then allocated patients to the second closest facility and there was also insufficient capacity in some cases, so some patients were allocated to the third closest facility. Such allocation challenges could be eliminated with the use of the DFNDP model since it includes the constraint that all patients are allocated to the closest facility. The results show that with the existing facilities, patients would experience a mean travel time of 23.50 minutes with standard deviation of 18.62 minutes. 53 patients would need to travel more than 45 minutes to get to a dialysis facility. In the worst case, a patient would need to travel 102 minutes to a dialysis facility.

We noticed that the home dialysis proportion with existing facilities is 18.9% whereas the home dialysis proportion in the DFNDP results is typically more than 30%. Achieving such home dialysis proportions may or may not be possible in practice. Our patient survey results showed that some HD patients would switch to home dialysis if training is available in Yarmouth. If the need to travel to Halifax for dialysis training remains, then the home dialysis proportion will be less. It is also possible that some survey participants may prefer home dialysis, but nephrologists can recommend facility-based dialysis for clinical reasons.

Table 4–4: DFNDP Unadjusted Test Results

	Objective Value (Normalized to base case)	Home Dialysis	$max(t_{ij})$ (minutes)	$mean(t_{ij})$ (minutes)	$stdev(t_{ij})$ (minutes)	Number of patients with $t_{ij} > T$
Existing Facilities	438,647,290,079,851,000,000	18.9%	102	23.50	18.62	53
$T = 30$	3,269,019	31.7%	54	16.51	11.54	58
$T = 45$ (base)	1	31.2%	54	16.53	11.40	8
$T = 60$	0.0000003	30.5%	54	17.22	11.83	0
$T = 75$	0.0000000	39.1%	59	18.71	13.35	0
$T = 90$	0.0000000	42.5%	71	18.24	15.95	0
$U = 30$	1.8329138	32.9%	55	16.28	11.22	7
$U = 60$	1.0000031	29.9%	54	16.77	11.63	8
$U = 75$	1.0000000	31.2%	54	16.38	11.43	8
$U = 90$	1.0000000	31.2%	54	16.50	11.39	8
budget + 2%	0.0043054	34.1%	48	15.05	10.30	4
budget + 4%	0.0017928	29.3%	48	16.10	10.31	2
budget + 6%	0.0013142	29.4%	48	15.42	9.75	1
budget + 8%	0.0012528	29.4%	48	15.62	9.50	1
budget + 10%	0.0012221	30.5%	48	5.28	9.38	1

Furthermore, it is possible that since the survey participants all live in rural areas within the NSHA Western Zone, they may find home dialysis relatively more attractive and be more willing to participate in home dialysis compared with dialysis patients from other parts of the province. Therefore, we ran additional *conservative* tests. We increased the proportion of dialysis patients who always go to an in-centre or satellite facility regardless of travel time by 25%. We report the remaining results and figures with this adjustment and still demonstrate the possibility for considerable improvements with our DFNDP model compared to existing facilities. We report the *conservative* test results in Table 4-5.

In the conservative case of the DFNDP model, we still find a lower mean travel time of approximately 20 minutes with less variability and significantly lower worst case of 73 to 74 minutes compared to existing facilities. These results are obtained with the same budget as found with existing facilities. As the budget is increased, the extent that mean travel times and variability are further reduced are shown in the table. We observe that the maximum travel time can be reduced to less than 60 minutes with a 4% increase in budget. We also observe that in the base case, 45 patients would need to travel more than 45 minutes, and

Table 4–5: DFNDP Conservative Test Results

	Objective Value (Normalized to base case)	Home Dialysis	$max(t_{ij})$ (minutes)	$mean(t_{ij})$ (minutes)	$stdev(t_{ij})$ (minutes)	Number of patients with $t_{ij} > T$
Existing Facilities	2,451,079,308,408	18.9%	102	23.50	18.62	53
$T = 30$	3,269,019	22.9%	73	20.15	15.46	103
$T = 45$ (base)	1	21.3%	73	20.37	15.50	45
$T = 60$	0.0000003	22.0%	73	20.24	15.42	6
$T = 75$	0.0000000	21.2%	73	20.51	15.41	0
$T = 90$	0.0000000	21.0%	74	19.92	15.46	0
$U = 30$	1.0000000	22.9%	73	20.04	15.46	45
$U = 60$	1.0000000	22.9%	73	20.04	15.45	45
$U = 75$	1.0000000	22.4%	73	20.11	15.40	45
$U = 90$	1.0000000	22.2%	73	20.15	15.41	45
budget + 2%	0.0102206	21.2%	69	19.35	14.41	39
budget + 4%	0.0000005	20.1%	59	17.63	12.65	23
budget + 6%	0.0000001	20.0%	57	17.09	11.96	23
budget + 8%	0.0000001	20.3%	57	16.53	11.18	9
budget + 10%	0.0000001	20.0%	57	15.80	10.15	6

increasing the budget could result in a dialysis facility network where fewer patients would need to travel more than 45 minutes. For example, with a 4% increase in budget, 23 patients would need to travel more than 45 minutes, whereas with an 8% increase in budget, 9 patients would need to travel more than 45 minutes with the optimal network of dialysis facilities.

A map representing the facility location decisions for the optimal solution of the base case is provided in Figure 4-8. The locations numbered above 14 represent potential facility locations that have never had dialysis facilities. Within the Western Zone, the optimal solution includes the existing satellite facility in Liverpool (6), plus satellite facilities in Digby (27) and Kentville (24). While Kentville (24) seems to replace Berwick (7) within the Annapolis Valley, Digby seems to be an important location that is missing from the existing dialysis facility network. The optimal solution also includes an additional facility to accomodate rural areas in the Eastern Zone (32) but fewer satellite facilities in the Northern Zone. Interstingly, the optimal solution also includes additional satellite facilities at (or near) the East Coast Forensic Hospital in Dartmouth (30) and at the Musquodoboit Valley

Memorial Hospital in Middle Musquodoboit (31) within the Central Zone. Therefore, the solution seems to accomodate improvements for patients from both rural and urban areas.

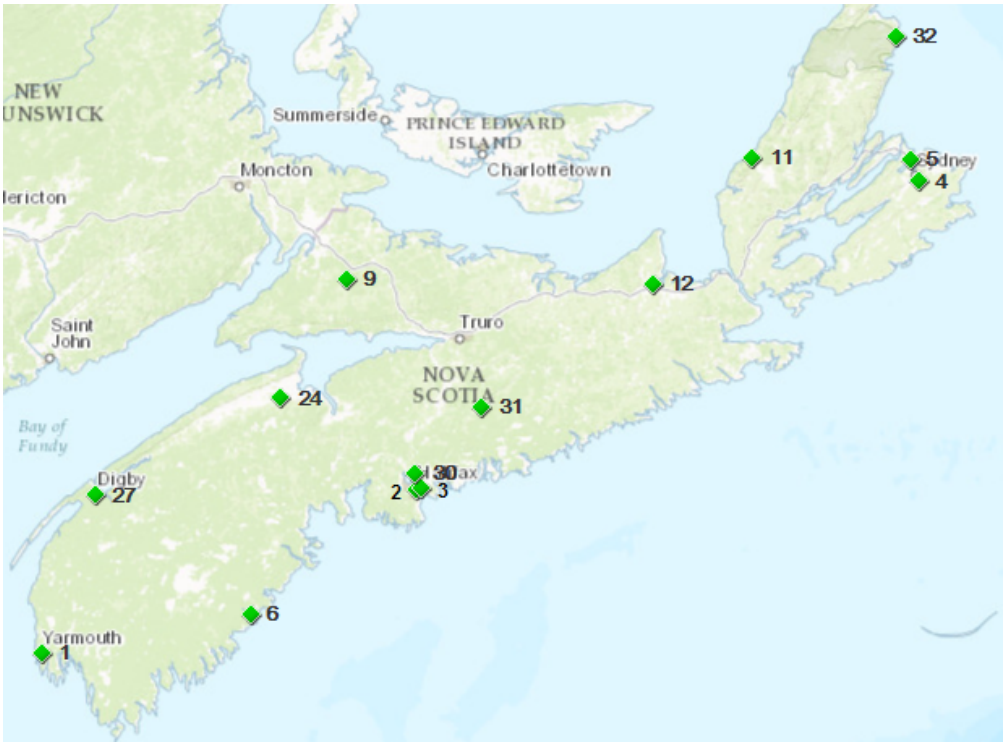

Figure 4–8: Facility Locations for Optimal Solution (base case)

Capacity decisions for In-Centre Facilities are provided in Figure 4-9. The optimal solution includes expanding in-centre capacity at the Dartmouth General Hospital (3) and leaving capacity at existing minimum levels for the other in-centre facilities. If the minimum in-centre capacity requirement constraints are relaxed, the optimal solution would have lower capacity in some of the in-centre facility locations. However, some or all of this additional capacity may be required for acute care purposes.

The parametric analysis on T and U has little impact on the facility location decisions, but some impact on capacity decisions for satellite facilities. Increasing T results in some moderate changes in the optimal capacity decisions for the satellite facility locations as shown in Figure 4-10. The only facility location change occurs for facility (24) which is replaced by nearby facility (7) when $T > 60$. Similarly, the parametric analysis on U has the same

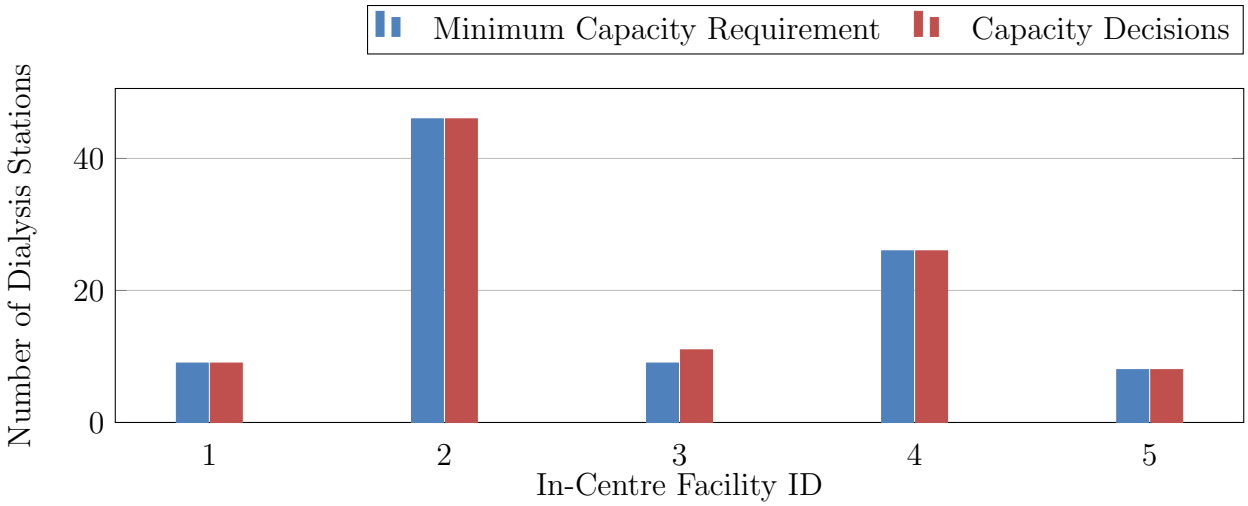

Figure 4–9: In-Centre Facility Location and Capacity Decisions, tests with B =base budget

optimal facility location decisions but there are some changes to capacity decisions for the satellite facilities as shown in Figure 4-11.

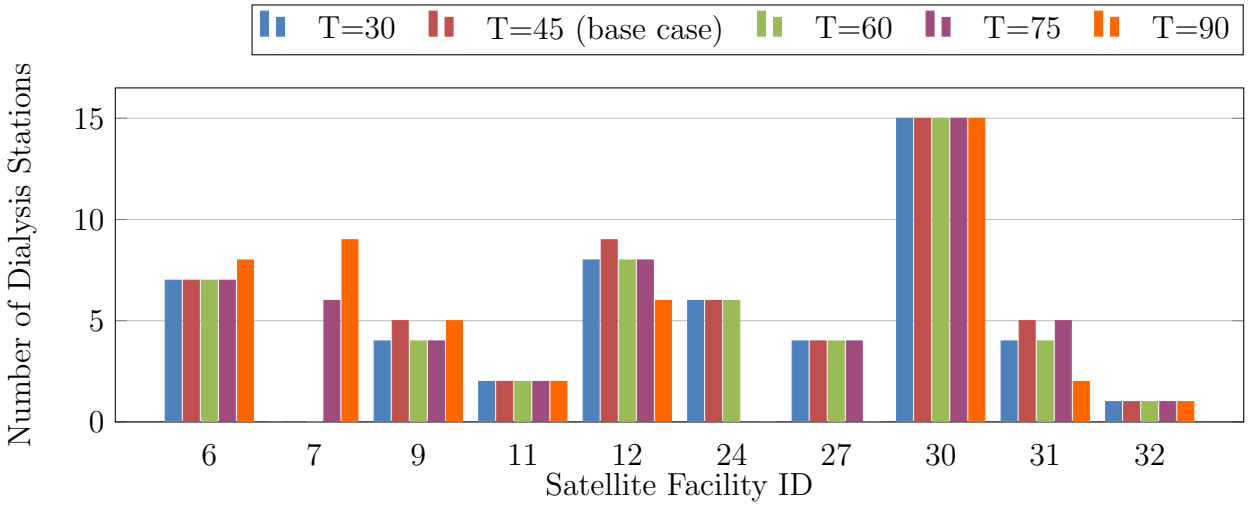

Figure 4–10: Satellite Facility Location and Capacity Decisions, $T = 30, 45, 60, 75, 90$, $U = 45$, B =base budget

When we increase the budget parameter we find that the optimal facility network changes as shown in the map figures 4-12 to 4-16 for increases of 2%, 4%, 6%, 8%, and 10% respectively. Increases to the available budget would result in the ability for additional facilities to be built. Among the facilities that are added to the dialysis facility network are the existing facilities, with the exception of facility (14) which never appears in the optimal

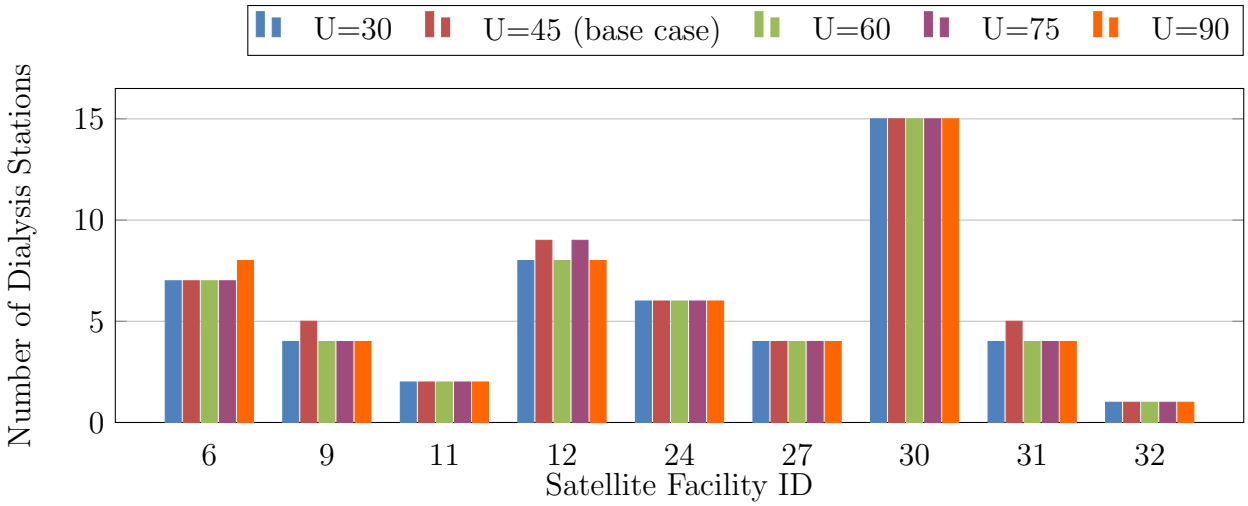

Figure 4–11: Satellite Facility Location and Capacity Decisions, $T = 45$, $U = 30, 45, 60, 75, 90$, B =base budget

solution. Interestingly, facility (14) was closed but probably would never have been opened if the DFNDP model was used as a planning tool in Nova Scotia.

4.7 Conclusion

Planning a network of dialysis facilities in a budget constrained health care system can be challenging. Adding the fact that patients can choose to go to a facility or perform dialysis at home makes the dialysis facility location and capacity management problem more challenging. With these constraints, it may seem impossible to provide reasonable travel times to patients in rural areas. However, with the DFNDP model, we provide a way for the design of a network of dialysis facilities that is the best possible network given the system constraints. The proposed model does not use hard constraints on maximum or average travel time, and as a result we always find an optimal solution with the proposed DFNDP model.

Our application of the DFNDP model to the province of Nova Scotia, Canada shows significant improvements compared to the existing facility network. We find that while the DFNDP ensures that sufficient capacity exists so that all patients visit the closest facility, the existing facility network results in some patients having to be allocated to their second closest or third closest facility. Overall, we find that the DFNDP results in reduction in

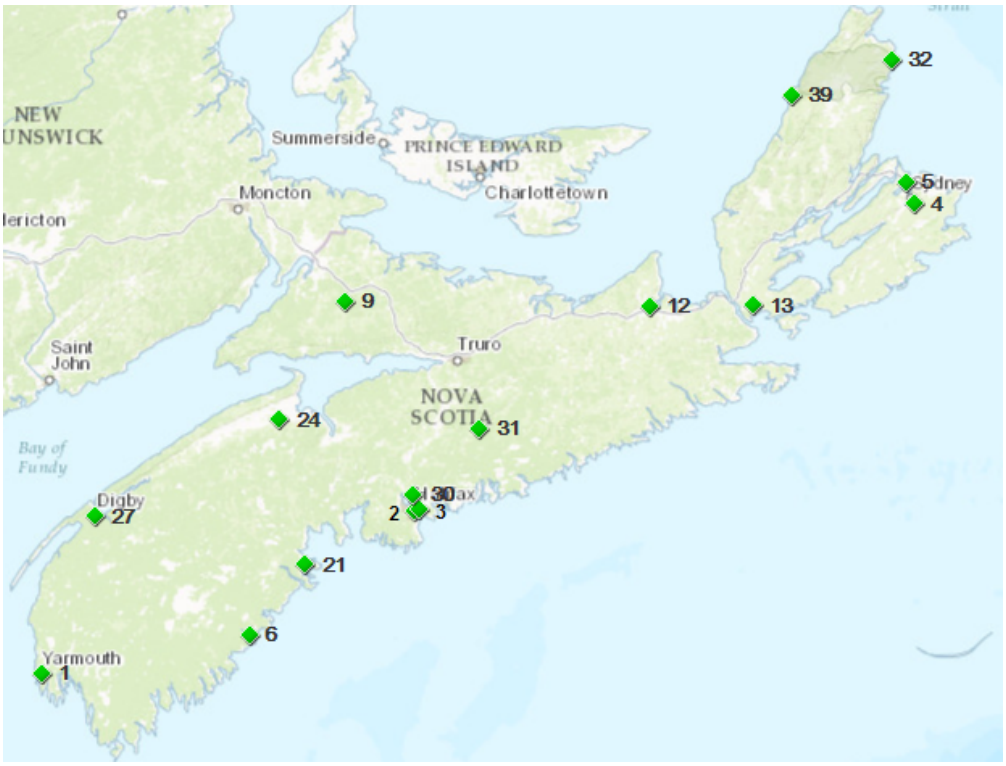

Figure 4–12: Facility Locations for Optimal Solution (base budget + 2%)

maximum and mean travel times and less variability in travel times compared to existing facilities. We found that considerable improvements are possible with the same budget. Furthermore, we illustrate the best location and capacity expansion decisions if additional budget is available. This would be particularly helpful to quantify the benefit of additional funding to support improvements to the dialysis facility network.

Although the case study considered the situation of planning satellite dialysis facilities before there were any satellite facilities, the model can be used for other purposes as well. For example, the model can be used to determine the location and capacity for a new satellite facility to be added to an existing network of satellite dialysis facilities. One would simply need to add constraints to the model for satellite facilities as we have done for the existing in-centre facilities in the case study presented within this chapter.

We also note that the static model proposed in this chapter does not account for the fact that the number of patients changes over time and that patients can switch dialysis modes.

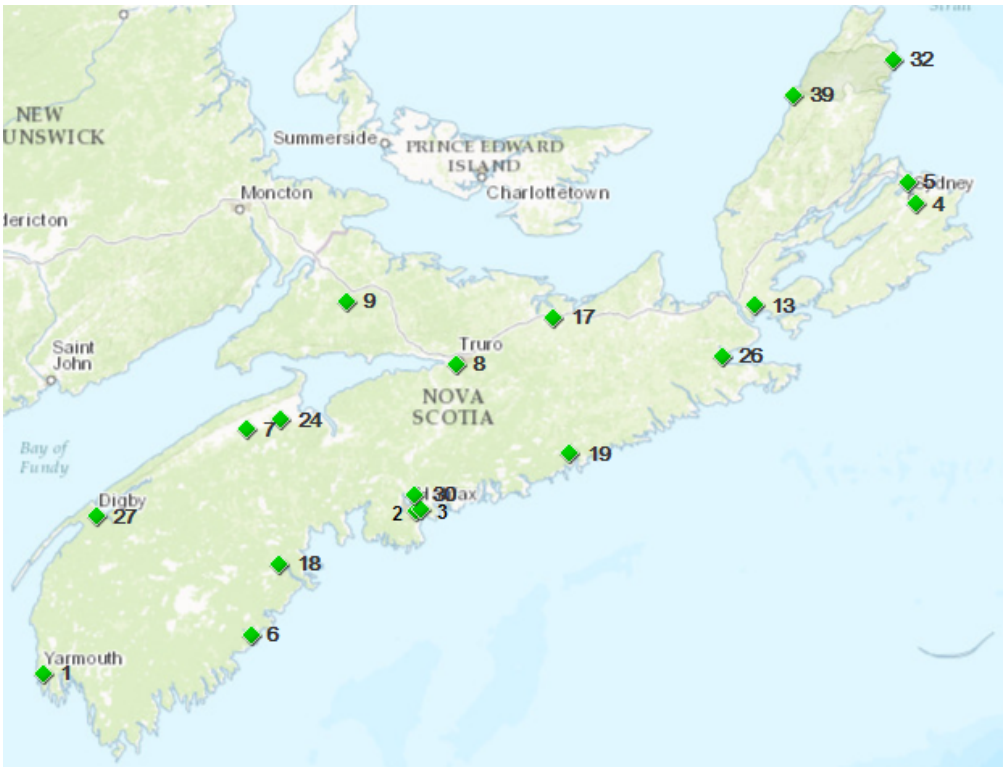

Figure 4–13: Facility Locations for Optimal Solution (base budget + 4%)

Our future research includes the development of a queueing network model to account for the changes in requirements due to switching dialysis modes or for new patients based on ESRD incidence and reductions in dialysis needs due to transplant or death. However, in the mean time, the current model can be used if run multiple times to test out various scenarios with different patient sets. For example, a set of patient addresses could be generated as we have done in the case study, and then some of those patients could be randomly removed from the set and additional patients could be added. For each generated potential patient set, the DFNDP could be run to gain an understanding on the impact that such changes would have on the optimal dialysis facility network.

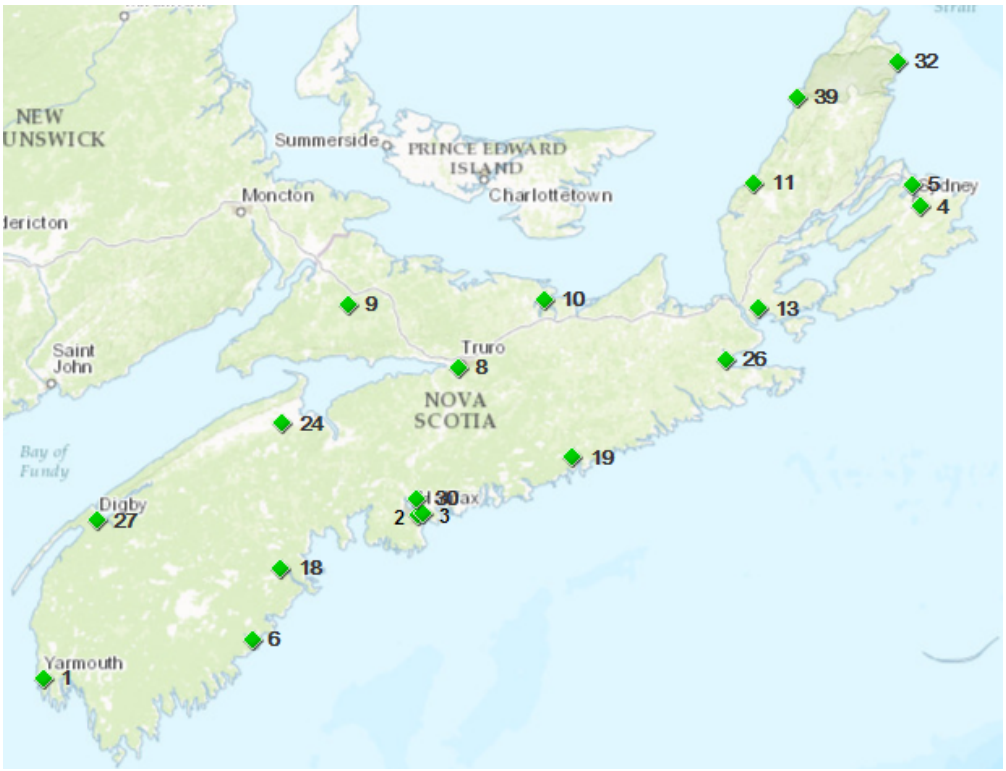

Figure 4–14: Facility Locations for Optimal Solution (base budget + 6%)

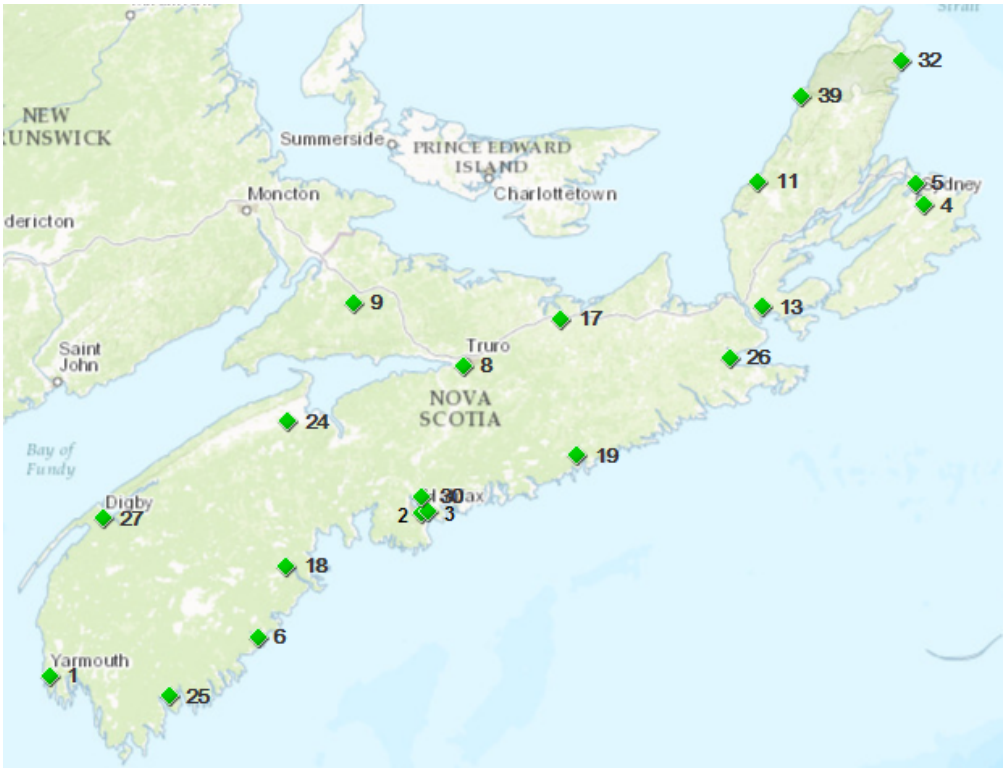

Figure 4–15: Facility Locations for Optimal Solution (base budget + 8%)

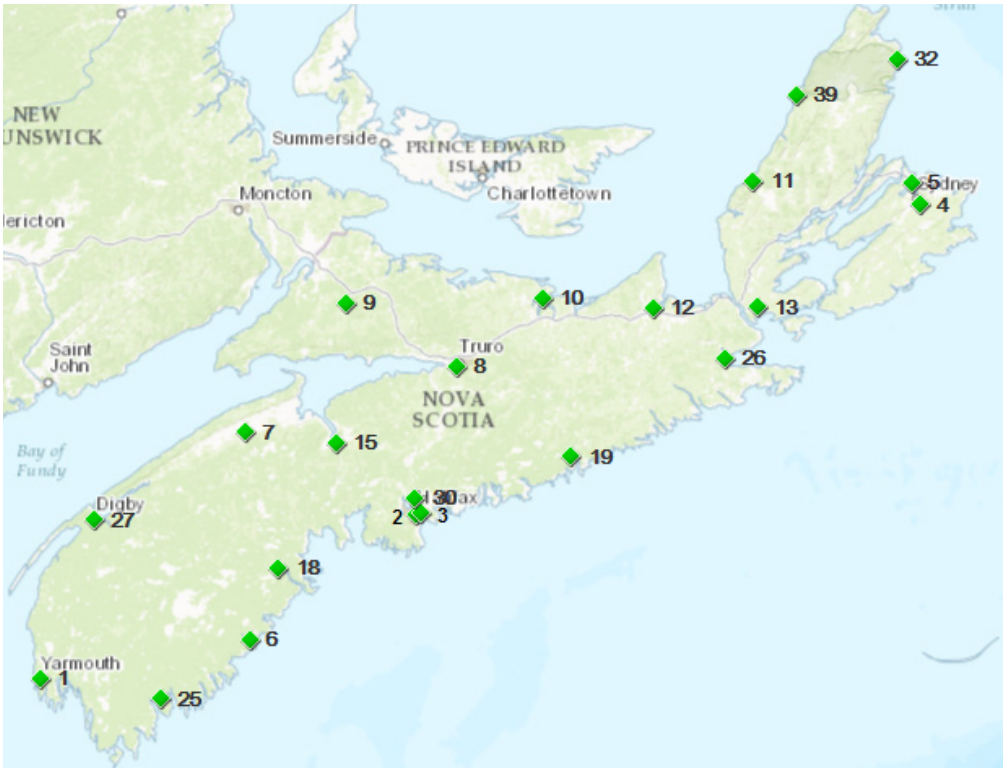

Figure 4–16: Facility Locations for Optimal Solution (base budget + 10%)

Chapter 5
Concluding Remarks and Future Research

Specialist Care is an understudied area of healthcare operations research, and rural healthcare has also received little research attention. In the research projects presented in this thesis, operations management challenges for specialist care in rural areas have been considered for both acute and chronic care processes. Studying the acute care topic of ED to ward patient flow from the specialist's perspective helps bring insight into the interdepartmental challenges of the ED boarding problem beyond bed-based capacity management. Specialists workload includes the challenge of managing the inflow of potential inpatients through ED consultations, taking care of inpatients after admission to hospital wards, and managing the outflow of inpatients through inpatient discharges. This interdepartmental process is challenging in any setting, but the rural case is more complex for the Internists who manage ICU care on top of Internal Medicine care responsibilites. This thesis also considers the challenging chronic dialysis facility network design problem, where long travel times impact patient welfare, particular for those who live in rural areas. The DFNDP model proposed can help identify the best network of dialysis facilities with consideration for budget and capacity management constraints as well as patient preferences for facility or home dialysis.

For the first project, I obtained the consent of every Internist in two regional hospitals to participate and observed their work on call in the hospital. These collaborations allowed for successful data collection efforts from both Internists and hospital information systems analysts. Our proposed dynamic programming framework for Specialist Care includes both Single Role and Dual Role models. We avoided the need for approximation methods due to our modeling approach which includes novel elements and appropriate assumptions that we observed in practice. As a result, we were able to illustrate the application of our models

with two corresponding case studies. We found that an early inpatient discharge strategy (i.e. discharge patients by 11:00AM) is generally not a good strategy. Instead, specialists should sometimes give priority for ED consultations and other times give priority to inpatient discharges. We found that optimal policies include boarding thresholds and end-of-horizon effects.

With the second project, I developed an optimization model for dialysis services planning that incorporates patient choice for dialysis mode. A feasible solution is always possible with the model formulation so that patient welfare can be maximized to the extent possible given budget and capacity constraints and patient preferences. Our proposed model is a mixed integer program (MIP) that can be solved with a standard commercial MIP solver such as CPLEX. To the best of our knowledge, the proposed model is the first optimization model for dialysis facility network design that incorporates patient choice for facility-based or home dialysis. To apply the model to the province of Nova Scotia, I used data from the patient survey, cost information that we obtained from Nova Scotia Renal Program management, and I generated realistic patient location data using public information including census data from Statistics Canada. We found that with the same budget as the cost to build and operate the existing facility network, considerable reductions of maximum and mean travel times are possible along with less variability. Furthermore, I demonstrated how the model can also help determing the best use of additional funding by determining optimal location and capacity expansion decisions.

This thesis sets the stage for further healthcare operations research in both rural and urban contexts. The literature review on Emergency Care in Chapter 2 revealed that there is a lack of Emergency Care studies in rural hospitals. While many aspects of medical procedures are standard in rural and urban hospitals alike, further research in rural EDs is needed to determine exactly how similar these processes are compared to urban EDs. Chapter 2 also identified that the ED operations management literature includes some studies that incorporate unique features of ED patient flow such as triage and lab tests, but ED

consultations from specialists had not been studied in the operations management literature before the project undertaken for Chapter 3 of this thesis. It will be interesting to determine if the proposed Single Role model is appropriate for urban hospitals. The Single Role case that we found in one of our rural study hospitals is similar to the situation in urban hospitals. However, in the case of Internal Medicine, there is usually a team of Internists working on call in urban hospitals rather than a single Internist as we observed in rural hospitals. In urban hospitals where one specialist acts as a lead specialist responsible for the decisions for ED consultation and inpatient discharge timing, then the proposed Single Role model should be adequate. However, when multiple specialists are working on call at the same time, there is opportunity to consider different management approaches for Specialist Care, which to our knowledge, has not yet been studied.

From a methodological perspective, our future research includes examining continuous time models for the workflow decisions of specialists rather than the discrete time models proposed in Chapter 3. We observe a quasi-birth-and-death (QBD) structure which may be best analyzed in continuous time as an extension to the models proposed in this thesis. It will be interesting to know if the models presented in this thesis and/or extensions completed in future research are applicable to both rural and urban hospitals. If there are significant differences, further study on the workflow decisions of specialists in urban hospitals will also be needed.

As there are many different medical specialties required to support the needs of different chronic conditions, great opportunity exists to develop a variety of new models to help manage chronic care. In this thesis, we focus on dialysis for patients with ESRD. Note however, that chronic conditions usually require care from specialists, so additional studies on the role of specialists in supporting other chronic conditions is another possible research direction.

In the case of the Dialysis Facility Network Design Probelm (DFNDP), the model proposed in Chapter 4 does not consider the fact that patients can switch back and forth from

home dialysis to facility-based dialysis. We also considered a static model which does not account for the changes in the number and location of dialysis patients over time. Although the proposed DFNDP model could be run multiple times to consider different numbers of patients at different locations, our future research may tackle additional complexity in planning a network of dialysis facilities. In particular, we have started to develop a queueing network model to account for the changes in requirements due to switching dialysis modes or for new patients based on ESRD incidence and reductions in dialysis needs due to transplant or death.

Overall, although general models can be developed for some Specialist Care processes, there are also specific opportunities for Applied Operations Research in specific areas including Cardiology, Gastroenterology, and Nephrology. I also look forward to further studies on Rural and Urban Health Services Management, including Emergency Department Crowding and other important research topics.

Appendix - Dialysis Patient Survey

Dialysis Patient Survey

STUDY TITLE:	Facility Network Design for Dialysis
PRINCIPAL INVESTIGATOR:	Michael G. Klein, Ph.D. candidate, McGill University, Desautels Faculty of Management, 1001 Sherbrooke St. W. Montreal, QC H3A 1G5 (902) 482-1300 michael.klein2@mail.mcgill.ca
STUDY SPONSOR:	McGill University, Desautels Faculty of Management
FUNDER:	This study is being funded by the NSERC CREATE Program on Healthcare Operations and Information Management

---

The Nova Scotia Health Authority (NSHA) Research Ethics Board (REB) requires that the informed consent form is signed before completing this survey.

Please ask the research team to clarify anything you do not understand or would like to know more about.

Thank you very much for agreeing to participate in this study.

NAME: _____ DATE: _____
(DD / MM / YYYY)

ADDRESS: _____

_____

AGE: _____ GENDER: ☐ MALE ☐ FEMALE

WORK STATUS: ☐ EMPLOYED ☐ RETIRED ☐ OTHER: _____

1A. Which of the following choices were presented to you as possible methods of treatment for your kidney failure? (check all that apply)

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility (check all that apply)

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at a satellite dialysis facility (check all that apply)

   ☐ Inverness ☐ Antigonish ☐ Pictou ☐ Springhill ☐ Liverpool

   ☐ Cleveland ☐ Sherbrooke ☐ Truro ☐ Berwick

3. ☐ Hemodialysis (HD) at home, after training in Halifax
4. ☐ Peritoneal Dialysis (PD) at home, after training in Halifax
5. ☐ Kidney Transplant

_____

2A. What form of dialysis are you currently receiving for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at a satellite dialysis facility

   ☐ Inverness ☐ Antigonish ☐ Pictou ☐ Springhill ☐ Liverpool

   ☐ Cleveland ☐ Sherbrooke ☐ Truro ☐ Berwick

3. ☐ Hemodialysis (HD) at home, after training in Halifax
4. ☐ Peritoneal Dialysis (PD) at home, after training in Halifax

2B. If you are receiving hemodialysis (HD) at an in-centre or satellite dialysis facility, you do not use home dialysis because:

_____

_____

_____

_____

2C. Is your current form of treatment the same as the one you started on?

☐ YES ☐ NO

2D. If NO, the type of treatment was changed from _____

because: _____

_____

_____

_____

Research Ethics Board

3A. If the closest dialysis facility was a satellite facility within a **15 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **15 minute** drive)
3. ☐ Hemodialysis (HD) at home, after training in Halifax
4. ☐ Peritoneal Dialysis (PD) at home, after training in Halifax

3B. If the closest dialysis facility was a satellite facility within a **30 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **30 minute** drive)
3. ☐ Hemodialysis (HD) at home, after training in Halifax
4. ☐ Peritoneal Dialysis (PD) at home, after training in Halifax

3C. If the closest dialysis facility was a satellite facility within a **45 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **45 minute** drive)
3. ☐ Hemodialysis (HD) at home, after training in Halifax
4. ☐ Peritoneal Dialysis (PD) at home, after training in Halifax

3D. If the closest dialysis facility was a satellite facility within a **60 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **60 minute** drive)
3. ☐ Hemodialysis (HD) at home, after training in Halifax
4. ☐ Peritoneal Dialysis (PD) at home, after training in Halifax

3E. If the closest dialysis facility was a satellite facility within a **90 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **90 minute** drive)
3. ☐ Hemodialysis (HD) at home, after training in Halifax
4. ☐ Peritoneal Dialysis (PD) at home, after training in Halifax

---

4A. If training for home dialysis was available at all in-centre dialysis facilities and the closest dialysis facility was a satellite facility within a **15 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **15 minute** drive)
3. ☐ Hemodialysis (HD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

4. ☐ Peritoneal Dialysis (PD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

Research Ethics Board

4B. If training for home dialysis was available at all in-centre dialysis facilities and the closest dialysis facility was a satellite facility within a **30 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **30 minute** drive)
3. ☐ Hemodialysis (HD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

4. ☐ Peritoneal Dialysis (PD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

4C. If training for home dialysis was available at all in-centre dialysis facilities and the closest dialysis facility was a satellite facility within a **45 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **45 minute** drive)
3. ☐ Hemodialysis (HD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

4. ☐ Peritoneal Dialysis (PD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

Research Ethics Board

4D. If training for home dialysis was available at all in-centre dialysis facilities and the closest dialysis facility was a satellite facility within a **60 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **60 minute** drive)
3. ☐ Hemodialysis (HD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

4. ☐ Peritoneal Dialysis (PD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

4E. If training for home dialysis was available at all in-centre dialysis facilities and the closest dialysis facility was a satellite facility within a **90 minute** drive from your home, what form of treatment would you choose for your kidney failure?

1. ☐ Hemodialysis (HD) at an in-centre dialysis facility

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

2. ☐ Hemodialysis (HD) at the closest satellite dialysis facility (within a **90 minute** drive)
3. ☐ Hemodialysis (HD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

4. ☐ Peritoneal Dialysis (PD) at home, **after training at an in-centre dialysis facility**

   ☐ Sydney ☐ North Sydney ☐ Halifax ☐ Dartmouth ☐ Yarmouth

Thank you very much for participating in this study.

References

Abo-Hamad, Waleed, Amr Arisha. 2013. Simulation-based framework to improve patient experience in an emergency department. *European Journal of Operational Research* **224**(1) 154–166.

Aboolian, Robert, Oded Berman, Vedat Verter. 2015. Maximal Accessibility Network Design in the Public Sector. *Transportation Science* .

Adams, E. Kathleen, Robert Houchens, George E. Wright, James Robbins. 1991. Predicting hospital choice for rural Medicare beneficiaries: the role of severity of illness. *Health Services Research* **26**(5) 583.

Adler, Nancy E., David H. Rehkopf. 2008. US disparities in health: descriptions, causes, and mechanisms. *Annu. Rev. Public Health* **29** 235–252.

Ahmed, Mohamed A., Talal M. Alkhamis. 2009. Simulation optimization for an emergency department healthcare unit in Kuwait. *European Journal of Operational Research* **198**(3) 936–942.

Alagoz, Oguzhan, Lisa M. Maillart, Andrew J. Schaefer, Mark S. Roberts. 2004. The Optimal Timing of Living-Donor Liver Transplantation. *Management Science* **50**(10) 1420–1430.

Alagoz, Oguzhan, Lisa M. Maillart, Andrew J. Schaefer, Mark S. Roberts. 2007. Choosing Among Living-Donor and Cadaveric Livers. *Management Science* **53**(11) 1702–1715.

Allon, Gad, Sarang Deo, Wuqin Lin. 2013. The Impact of Size and Occupancy of Hospital on the Extent of Ambulance Diversion: Theory and Evidence. *Operations Research* .

Almehdawe, Eman, Beth Jewkes, Qi-Ming He. 2013. A Markovian queueing model for ambulance offload delays. *European Journal of Operational Research* **226**(3) 602–614.

Arcury, Thomas A., Wilbert M. Gesler, John S. Preisser, Jill Sherman, John Spencer, Jamie Perin. 2005. The effects of geography and spatial behavior on health care utilization among the residents of a rural region. *Health services research* **40**(1) 135–156.

Armony, Mor, Amy R. Ward. 2010. Fair Dynamic Routing in Large-Scale Heterogeneous-Server Systems. *Operations Research* **58**(3) 624–637.

Asplin, Brent R., Thomas J. Flottemesch, Bradley D. Gordon. 2006. Developing Models for Patient Flow and Daily Surge Capacity Research. *Academic Emergency Medicine* **13**(11) 1109–1113.

Asplin, Brent R., David J. Magid, Karin V. Rhodes, Leif I. Solberg, Nicole Lurie, Carlos A. Camargo Jr. 2003. A conceptual model of emergency department crowding. *Annals of Emergency Medicine* **42**(2) 173–180.

Baernholdt, Marianne, Barbara A. Mark. 2009. The nurse work environment, job satisfaction and turnover rates in rural and urban nursing units. *Journal of nursing management* **17**(8) 994–1001.

Bair, Aaron E., Wheyming T. Song, Yi-Chun Chen, Beth A. Morris. 2010. The Impact of Inpatient Boarding on ED Efficiency: A Discrete-Event Simulation Study. *Journal of Medical Systems* **34**(5) 919–929.

Beaulieu, Huguette, Jacques A. Ferland, Bernard Gendron, Philippe Michelon. 2000. A mathematical programming approach for scheduling physicians in the emergency room. *Health Care Management Science* **3**(3) 193–200.

Bernstein, Steven L., Dominik Aronsky, Reena Duseja, Stephen Epstein, Dan Handel, Ula Hwang, Melissa McCarthy, K. John McConnell, Jesse M. Pines, Niels Rathlev, Robert Schafermeyer, Frank Zwemer, Michael Schull, Brent R. Asplin, Emergency Department Crowding Task Force Society for Academic Emergency Medicine. 2009. The Effect of Emergency Department Crowding on Clinically Oriented Outcomes. *Academic Emergency Medicine* **16**(1) 1–10.

Bernstein, Steven L., Vinu Verghese, Winifred Leung, Anne T. Lunney, Ivelisse Perez. 2003. Development and Validation of a New Index to Measure Emergency Department Crowding. *Academic Emergency Medicine* **10**(9) 938–942.

Beveridge, Robert, James Ducharme, Laurie Janes, Serge Beaulieu, Stephen Walter. 1999. Reliability of the Canadian Emergency Department Triage and Acuity Scale: Interrater Agreement. *Annals of Emergency Medicine* **34**(2) 155–159.

Blagg, Christopher R. 1999. The early years of chronic dialysis: The Seattle contribution. *American journal of nephrology* **19**(2) 350–354.

Blagg, Christopher R. 2007. The early history of dialysis for chronic renal failure in the United States: a view from Seattle. *American Journal of Kidney Diseases* **49**(3) 482–496.

Bosanac, Edward M., Rosalind C. Parkinson, David S. Hall. 1976. Geographic access to hospital care: a 30-minute travel time standard. *Medical Care* **14**(7) 616–623.

Brescia MJ, Cimino JE, Appel K, Hurwich BJ. 1966. Chronic hemodialysis using venipuncture and a surgically created arteriovenous fistula. *The New England journal of medicine* **275**(20) 1089–92.

Buykx, Penny, John Humphreys, John Wakerman, Dennis Pashen. 2010. Systematic review of effective retention incentives for health workers in rural and remote areas: Towards evidence-based policy. *Australian Journal of Rural Health* **18**(3) 102–109.

Carter, Michael W., Sophie D. Lapierre. 2001. Scheduling Emergency Room Physicians. *Health Care Management Science* **4**(4) 347–360.

Ceglowski, R, L Churilov, J Wasserthiel. 2006. Combining Data Mining and Discrete Event Simulation for a value-added view of a hospital emergency department. *J Oper Res Soc* **58**(2) 246–254.

Chao, Chia-Ter, Chun-Fu Lai, Jenq-Wen Huang, Chih-Kang Chiang, Sheng-Jen Huang. 2015. Association of increased travel distance to dialysis units with the risk of anemia in rural chronic hemodialysis elderly. *Hemodialysis International* **19**(1) 44–53.

Cho, Suck Ju, Jinwoo Jeong, Sangkyoon Han, Seokran Yeom, Sung Wook Park, Hyung Hoi Kim, Seong Youn Hwang. 2011. Decreased Emergency Department Length of Stay by Application of a Computerized Consultation Management System. *Academic Emergency Medicine* **18**(4) 398–402.

CIHI. 2015. 2015 CORR Report: Treatment of End-Stage Organ Failure in Canada, 2004 to 2013. Tech. rep.

Cochran, Jeffery K., Kevin T. Roche. 2009. A multi-class queuing network analysis methodology for improving hospital emergency department performance. *Computers & Operations Research* **36**(5) 1497–1512.

Congdon, Peter. 2001. The Development of Gravity Models for Hospital Patient Flows under System Change: A Bayesian Modelling Approach. *Health Care Management Science* **4**(4) 289–304.

Connelly, Lloyd G., Aaron E. Bair. 2004. Discrete Event Simulation of Emergency Department Activity: A Platform for System-level Operations Research. *Academic Emergency Medicine* **11**(11) 1177–1185.

Cooper, Richard A. 2002. There's a shortage of specialists: is anyone listening? *Academic Medicine* **77**(8) 761–766.

Cunningham, Charles E., Ken Deal, Heather Rimas, Heather Campbell, Ann Russell, Jennifer Henderson, Anne Matheson, Blake Melnick. 2008. Using conjoint analysis to model the preferences of different patient segments for attributes of patient-centered care. *The Patient: Patient-Centered Outcomes Research* **1**(4) 317–330.

Dai, J. G., Pengyi Shi. 2014. A Two-Time-Scale Approach to Time-Varying Queues for Hospital Inpatient Flow Management. *Working Paper. Available at SSRN 2489533* .

Deo, Sarang, Itai Gurvich. 2011. Centralized vs. Decentralized Ambulance Diversion: A Network Perspective. *Management Science* **57**(7) 1300–1319.

DesMeules, Marie. 2006. *How healthy are rural Canadians?: An assessment of their health status and health determinants*. Canadian Institute for Health Information.

Dobson, Gregory, Hsiao-Hui Lee, Edieal Pinker. 2010. A Model of ICU Bumping. *Operations Research* .

Dobson, Gregory, Tolga Tezcan, Vera Tilson. 2013. Optimal Workflow Decisions for Investigators in Systems with Interruptions. *Management Science* .

Doucette, Keith. 2015. 'Living on overtime': Nova Scotia one of first hit by nurse shortage. *The Canadian Press* .

Eben-Chaime, Moshe, Joseph S. Pliskin. 1992. Incorporating patient travel times in decisions about size and location of dialysis facilities. *Medical Decision Making* **12**(1) 44–51.

Eberhardt, Mark S., Elsie R. Pamuk. 2004. The importance of place of residence: examining health in rural and nonrural areas. *American Journal of Public Health* **94**(10) 1682–1686.

Eckstein, Marc, Linda S Chan. 2004. The effect of emergency department crowding on paramedic ambulance availability. *Annals of Emergency Medicine* **43**(1) 100–105.

Eldabi, T., R. J. Paul, T. Young. 2007. Simulation modelling in healthcare: reviewing legacies and investigating futures. *Journal of the Operational Research Society* **58**(2) 262–270.

Epstein, Stephen K., Lu Tian. 2006. Development of an Emergency Department Work Score to Predict Ambulance Diversion. *Academic Emergency Medicine* **13**(4) 421–426.

Falvo, Thomas, Lance Grove, Ruth Stachura, David Vega, Rose Stike, Melissa Schlenker, William Zirkin. 2007. The Opportunity Loss of Boarding Admitted Patients in the Emergency Department. *Academic Emergency Medicine* **14**(4) 332–337.

Fee, Christopher, Ellen J. Weber, Carley A. Maak, Peter Bacchetti. 2007. Effect of Emergency Department Crowding on Time to Antibiotics in Patients Admitted With Community-Acquired Pneumonia. *Annals of Emergency Medicine* **50**(5) 501–509.e1.

Ferrand, Yann, Michael Magazine, Uday S. Rao, Todd F. Glass. 2011. Building Cyclic Schedules for Emergency Department Physicians. *Interfaces* **41**(6) 521–533.

FitzGerald, Gerard, George A Jelinek, Deborah Scott, Marie Frances Gerdtz. 2010. Emergency department triage revisited. *Emergency Medicine Journal* **27**(2) 86–92.

Fletcher, A., D. Halsall, S. Huxham, D. Worthington. 2007. The DH Accident and Emergency Department model: a national generic model used locally. *Journal of the Operational Research Society* **58**(12) 1554–1562.

Fletcher, Adrian, Dave Worthington. 2009. What is a generic hospital model?a comparison of generic and specific hospital models of emergency patient flows. *Health Care Management Science* **12**(4) 374–391.

Forster, Alan J., Ian Stiell, George Wells, Alexander J. Lee, Carl Van Walraven. 2003. The Effect of Hospital Occupancy on Emergency Department Length of Stay and Patient Disposition. *Academic Emergency Medicine* **10**(2) 127–133.

Gaba, David M., Steven K. Howard. 2002. Fatigue among Clinicians and the Safety of Patients. *New England Journal of Medicine* **347**(16) 1249–1255.

Geskey, Joseph M., Glenn Geeting, Cheri West, Christopher S. Hollenbeak. 2013. Improved Physician Consult Response Times in an Academic Emergency Department After Implementation of an InstitutionalGuideline. *The Journal of Emergency Medicine* **44**(5) 999–1006.

Gilboy, Nicki, Paula Tanabe, Debbie A. Travers. 2005. The Emergency Severity Index Version 4: Changes to ESI Level 1 and Pediatric Fever Criteria. *Journal of Emergency Nursing* **31**(4) 357–362.

Gorelick, Marc H., Kenneth Yen, Hyun J. Yun. 2005. The effect of in-room registration on emergency department length of stay. *Annals of Emergency Medicine* **45**(2) 128–133.

Graber-Naidich, Anna, Michael W. Carter, Vedat Verter. 2014. Primary care network development: the regulator's perspective. *Journal of the Operational Research Society* .

Graff, Louis G, Steve Wolf, Robert Dinwoodie, David Buono, David Mucci. 1993. Emergency physician workload: A time study. *Annals of Emergency Medicine* **22**(7) 1156–1163.

Grano, Melanie L. De, D. J. Medeiros, David Eitel. 2008. Accommodating individual preferences in nurse scheduling via auctions and optimization. *Health Care Management Science* **12**(3) 228–242.

Green, Linda V., Peter J. Kolesar, Ward Whitt. 2007. Coping with Time-Varying Demand When Setting Staffing Requirements for a Service System. *Production and Operations Management* **16**(1) 13–39.

Green, Linda V., Sergei Savin, Ben Wang. 2006a. Managing Patient Service in a Diagnostic Medical Facility. *Operations Research* .

Green, Linda V., Joo Soares, James F. Giglio, Robert A. Green. 2006b. Using Queueing Theory to Increase the Effectiveness of Emergency Department Provider Staffing. *Academic Emergency Medicine* **13**(1) 61–68.

Gurvich, Itay, Ward Whitt. 2009. Queue-and-Idleness-Ratio Controls in Many-Server Service Systems. *Math. Oper. Res.* **34**(2) 363–396.

Halfin, Shlomo, Ward Whitt. 1981. Heavy-Traffic Limits for Queues with Many Exponential Servers. *Operations Research* **29**(3) 567–588.

Hall, Susan A., Jay S. Kaufman, Thomas C. Ricketts. 2006. Defining urban and rural areas in US epidemiologic studies. *Journal of Urban Health* **83**(2) 162–175.

Hancock, Christine, Alan Steinbach, Thomas S. Nesbitt, Shelley R. Adler, Colette L. Auerswald. 2009. Why doctors choose small towns: a developmental model of rural physician recruitment and retention. *Social science & medicine* **69**(9) 1368–1376.

Handel, Daniel A., Joshua A. Hilton, Michael J. Ward, Elaine Rabin, Frank L. Zwemer, Jr, Jesse M. Pines. 2010. Emergency Department Throughput, Crowding, and Financial Outcomes for Hospitals. *Academic Emergency Medicine* **17**(8) 840–847.

Hart, L. Gary, Eric H. Larson, Denise M. Lishner. 2005. Rural definitions for health policy and research. *American Journal of Public Health* **95**(7) 1149.

Hartley, David. 2004. Rural health disparities, population health, and rural culture. *American Journal of Public Health* **94**(10) 1675–1678.

Holden, Richard J. 2011. Lean Thinking in Emergency Departments: A Critical Review. *Annals of Emergency Medicine* **57**(3) 265–278.

Holdgate, Anna, Jenny Morris, Margaret Fry, Milan Zecevic. 2007. Accuracy of triage nurses in predicting patient disposition. *Australasian Emergency Nursing Journal* **10**(4) 189–190.

Holroyd, Brian R., Michael J. Bullard, Karen Latoszek, Debbie Gordon, Sheri Allen, Siulin Tam, Sandra Blitz, Philip Yoon, Brian H. Rowe. 2007. Impact of a Triage Liaison Physician on Emergency Department Overcrowding and Throughput: A Randomized Controlled Trial. *Academic Emergency Medicine* **14**(8) 702–708.

Hoot, Nathan R., Dominik Aronsky. 2008. Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine* **52**(2) 126–136.e1.

Hoot, Nathan R., Larry J. LeBlanc, Ian Jones, Scott R. Levin, Chuan Zhou, Cynthia S. Gadd, Dominik Aronsky. 2008. Forecasting Emergency Department Crowding: A Discrete Event Simulation. *Annals of Emergency Medicine* **52**(2) 116–125.

Hoot, Nathan R., Chuan Zhou, Ian Jones, Dominik Aronsky. 2007. Measuring and Forecasting Emergency Department Crowding in Real Time. *Annals of Emergency Medicine* **49**(6) 747–755.

Horwitz, Leora I., Tannaz Moin, Harlan M. Krumholz, Lillian Wang, Elizabeth H. Bradley. 2008. Consequences of Inadequate Sign-out for Patient Care. *Archives of Internal Medicine* **168**(16) 1755–1760.

Huang, Junfei, Boaz Carmeli, Avishai Mandelbaum. 2015. Control of Patient Flow in Emergency Departments, or Multiclass Queues with Deadlines and Feedback. *Operations Research* .

Hwang, Ula, John Concato. 2004. Care in the Emergency Department: How Crowded Is Overcrowded? *Academic Emergency Medicine* **11**(10) 1097–1101.

Hwang, Ula, Lynne Richardson, Elayne Livote, Ben Harris, Natasha Spencer, R. Sean Morrison. 2008. Emergency Department Crowding and Decreased Quality of Pain Care. *Academic Emergency Medicine* **15**(12) 1248–1255.

Izady, Navid, Dave Worthington. 2012. Setting staffing requirements for time dependent queueing networks: The case of accident and emergency departments. *European Journal of Operational Research* **219**(3) 531–540.

Jang, Ji Yeon, Sang Do Shin, Eui Jung Lee, Chang Bae Park, Kyoung Jun Song, Adam J. Singer. 2013. Use of a Comprehensive Metabolic Panel Point-of-Care Test to Reduce Length of Stay in the Emergency Department: A Randomized Controlled Trial. *Annals of Emergency Medicine* **61**(2) 145–151.

Jones, Simon Andrew, Mark Patrick Joy, Jon Pearson. 2002. Forecasting Demand of Emergency Care. *Health Care Management Science* **5**(4) 297–305.

Kane, Robert L. 1969. Determination of health care priorities and expectations among rural consumers. *Health Services Research* **4**(2) 142.

Kendall, David G. 1953. Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics* **24**(3) 338–354.

Kennebeck, Stephanie Spellman, Nathan L. Timm, Eileen Murtagh Kurowski, Terri L. Byczkowski, Scott D. Reeves. 2011. The Association of Emergency Department Crowding and Time to Antibiotics in Febrile Neonates. *Academic Emergency Medicine* **18**(12) 1380–1385.

Khare, Rahul K., Emilie S. Powell, Gilles Reinhardt, Martin Lucenti. 2009. Adding More Beds to the Emergency Department or Reducing Admitted Patient Boarding Times: Which Has a More Significant Influence on Emergency Department Congestion? *Annals of Emergency Medicine* **53**(5) 575–585.e2.

King, Diane L, David I Ben-Tovim, Jane Bassham. 2006. Redesigning emergency department patient flows: Application of Lean Thinking to health care. *Emergency Medicine Australasia* **18**(4) 391–397.

Klarenbach, Scott, Braden Manns. 2009. Economic evaluation of dialysis therapies. *Seminars in nephrology*, vol. 29. Elsevier, 524–532.

Klein, Michael G., Gilles Reinhardt. 2012. Emergency Department Patient Flow Simulations Using Spreadsheets. *Simulation in Healthcare* **7**(1) 40–47.

Kolff, W. J., H. Th J. Berk, Nurse M. Welle, A. J. W. Ley, E. C. Dijk, J. Noordwijk. 1944. The Artificial Kidney: a dialyser with a great area. *Acta Medica Scandinavica* **117**(2) 121–134.

Kolff, Willem J. 1965. First Clinical Experience with the Artificial Kidney. *Annals of Internal Medicine* **62**(3) 608–619.

Komenda, Paul, Michael Copland, Jay Makwana, Ogdjenka Djurdjev, Manish M. Sood, Adeera Levin. 2010. The cost of starting and maintaining a large home hemodialysis program. *Kidney international* **77**(11) 1039–1045.

Lane, D. C., C. Monefeldt, J. V. Rosenhead. 2000. Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. *Journal of the Operational Research Society* **51**(5) 518–531.

Lee, Alec M. 1966. *Applied queueing theory*. Macmillan; St. Martin's P., London; Melbourne [etc.; New York.

Lee, Chris P., Glenn M. Chertow, Stefanos A. Zenios. 2008. Optimal Initiation and Management of Dialysis Therapy. *Operations Research* **56**(6) 1428–1449.

Lee, Helen, Braden Manns, Ken Taub, William A. Ghali, Stafford Dean, David Johnson, Cam Donaldson. 2002. Cost analysis of ongoing care of patients with end-stage renal disease: the

impact of dialysis modality and dialysis access. *American Journal of Kidney Diseases* **40**(3) 611–622.

Lee-Lewandrowski, Elizabeth, John Nichols, Elizabeth Van Cott, Ricky Grisson, Abner Louissaint, Theodore Benzer, Kent Lewandrowski. 2009. Implementation of a Rapid Whole Blood D-Dimer Test in the Emergency Department of an Urban Academic Medical Center Impact on ED Length of Stay and Ancillary Test Utilization. *American Journal of Clinical Pathology* **132**(3) 326–331.

Lynn, Stephan G, Arthur L Kellermann. 1991. Critical decision making: Managing the emergency department in an overcrowded hospital. *Annals of Emergency Medicine* **20**(3) 287–292.

Mandelbaum, Avishai, Petar Momcilovic, Yulia Tseytlin. 2012. On Fair Routing from Emergency Departments to Hospital Wards: QED Queues with Heterogeneous Servers. *Management Science* **58**(7) 1273–1291.

Martin-Misener, Ruth, Sandra M. Reilly, Ardene Robinson Vollman. 2010. Defining the role of primary health care nurse practitioners in rural Nova Scotia. *CJNR (Canadian Journal of Nursing Research)* **42**(2) 30–47.

Mason, Suzanne, Ellen J. Weber, Joanne Coster, Jennifer Freeman, Thomas Locker. 2012. Time Patients Spend in the Emergency Department: England's 4-Hour RuleA Case of Hitting the Target but Missing the Point? *Annals of Emergency Medicine* **59**(5) 341–349.

Mayhew, L., D. Smith. 2008. Using queuing theory to analyse the Governments 4-h completion time target in Accident and Emergency departments. *Health Care Management Science* **11**(1) 11–21.

McCarthy, Melissa L., Scott L. Zeger, Ru Ding, Scott R. Levin, Jeffrey S. Desmond, Jennifer Lee, Dominik Aronsky. 2009. Crowding Delays Treatment and Lengthens Emergency Department Length of Stay, Even Among High-Acuity Patients. *Annals of Emergency Medicine* **54**(4) 492–503.e4.

McConnell, K. John, Loren A. Johnson, Nadia Arab, Christopher F. Richards, Craig D. Newgard, Tina Edlund. 2007. The On-Call Crisis: A Statewide Assessment of the Costs of Providing On-Call Specialist Coverage. *Annals of Emergency Medicine* **49**(6) 727–733.e18.

McConnell, K. John, Christopher F. Richards, Mohamud Daya, Stephanie L. Bernell, Cody C. Weathers, Robert A. Lowe. 2005. Effect of Increased ICU Capacity on Emergency Department Length of Stay and Ambulance Diversion. *Annals of Emergency Medicine* **45**(5) 471–478.

Mcfarlane, Philip A., Andreas Pierratos, Donald A. Redelmeier. 2002. Cost savings of home nocturnal versus conventional in-center hemodialysis. *Kidney international* **62**(6) 2216–2222.

McHugh, Megan, Paula Tanabe, Mark McClelland, Rahul K. Khare. 2012. More Patients Are Triaged Using the Emergency Severity Index Than Any Other Triage Acuity System in the United States. *Academic Emergency Medicine* **19**(1) 106–109.

Mehrotra R, Chiu Y, Kalantar-Zadeh K, Bargman J, Vonesh E. 2011. Similar outcomes with hemodialysis and peritoneal dialysis in patients with end-stage renal disease. *Archives of Internal Medicine* **171**(2) 110–118.

Mills, Angela M., Frances S. Shofer, Esther H. Chen, Judd E. Hollander, Jesse M. Pines. 2009. The Association between Emergency Department Crowding and Analgesia Administration in Acute Abdominal Pain Patients. *Academic Emergency Medicine* **16**(7) 603–608.

Moist, Louise M., Jennifer L. Bragg-Gresham, Ronald L. Pisoni, Rajiv Saran, Takashi Akiba, Stefan H. Jacobson, Shunichi Fukuhara, Donna L. Mapes, Hugh C. Rayner, Akira Saito, others. 2008. Travel time to dialysis as a predictor of health-related quality of life, adherence, and mortality: the Dialysis Outcomes and Practice Patterns Study (DOPPS). *American Journal of Kidney Diseases* **51**(4) 641–650.

Monroe, Adele C., Thomas C. Ricketts, Lucy A. Savitz. 1991. *Cancer in rural versus urban populations: a review*. Health Services Research Center, University of North Carolina at Chapel Hill.

Murray, Robert P, Michael Leroux, Edward Sabga, Wes Palatnick, Louis Ludwig. 1999. Effect of point of care testing on length of stay in an adult emergency department. *The Journal of Emergency Medicine* **17**(5) 811–814.

Olshaker, Jonathan S., Niels K. Rathlev. 2006. Emergency Department overcrowding and ambulance diversion: The impact and potential solutions of extended boarding of admitted patients in the Emergency Department. *The Journal of Emergency Medicine* **30**(3) 351–356.

Panayiotopoulos, J.-C., G. Vassilacopoulos. 1984. Simulating hospital emergency departments queuing systems: (GI/G/m(t)) : (IHFF/N/). *European Journal of Operational Research* **18**(2) 250–258.

Parker, Barnett R., V. Srinivasan. 1976. A Consumer Preference Approach to the Planning of Rural Primary Health-Care Facilities. *Operations Research* **24**(5) 991.

Paul, Sharoda A., Madhu C. Reddy, Christopher J. DeFlitch. 2010. A Systematic Review of Simulation Studies Investigating Emergency Department Overcrowding. *Simulation-Transactions of the Society for Modeling and Simulation International* **86**(8-9) 559–571.

Pines, Jesse M., Robert J. Batt, Joshua A. Hilton, Christian Terwiesch. 2011. The Financial Consequences of Lost Demand and Reducing Boarding in Hospital Emergency Departments. *Annals of Emergency Medicine* **58**(4) 331–340.

Pines, Jesse M., Sanjay Iyer, Maureen Disbot, Judd E. Hollander, Frances S. Shofer, Elizabeth M. Datner. 2008. The Effect of Emergency Department Crowding on Patient Satisfaction for Admitted Patients. *Academic Emergency Medicine* **15**(9) 825–831.

Pines, Jesse M., A. Russell Localio, Judd E. Hollander, William G. Baxt, Hoi Lee, Carolyn Phillips, Joshua P. Metlay. 2007. The Impact of Emergency Department Crowding Measures on Time to Antibiotics for Patients With Community-Acquired Pneumonia. *Annals of Emergency Medicine* **50**(5) 510–516.

Pines, Jesse M., Charles V. Pollack, Deborah B. Diercks, Anna Marie Chang, Frances S. Shofer, Judd E. Hollander. 2009. The Association Between Emergency Department Crowding and Adverse Cardiovascular Outcomes in Patients with Chest Pain. *Academic Emergency Medicine* **16**(7) 617–625.

Pitts, Stephen R., Jesse M. Pines, Michael T. Handrigan, Arthur L. Kellermann. 2012. National Trends in Emergency Department Occupancy, 2001 to 2008: Effect of Inpatient Admissions Versus Emergency Department Practice Intensity. *Annals of Emergency Medicine* **60**(6) 679–686.e3.

Popovich, Robert P., Jack W. Moncrief, Karl D. Nolph, Ahad J. Ghods, Zbylut J. Twardowski, W. K. Pyle. 1978. Continuous Ambulatory Peritoneal Dialysis. *Annals of Internal Medicine*

88(4) 449–456.

Powell, Emilie S., Rahul K. Khare, Arjun K. Venkatesh, Ben D. Van Roo, James G. Adams, Gilles Reinhardt. 2012. The Relationship between Inpatient Discharge Timing and Emergency Department Boarding. *The Journal of Emergency Medicine* **42**(2) 186–196.

Quinton, Wayne, David Dillard, Belding H. Scribner. 1960. Cannulation of blood vessels for prolonged hemodialysis. *ASAIO Journal* **6**(1) 104–113.

Ramakrishnan, M., D. Sier, P. G. Taylor. 2005. A two-time-scale model for hospital patient flow. *IMA Journal of Management Mathematics* **16**(3) 197–215.

Reeder, Timothy J., Deeanna L. Burleson, Herbert G. Garrison. 2003. The Overcrowded Emergency Department: A Comparison of Staff Perceptions. *Academic Emergency Medicine* **10**(10) 1059–1064.

Reschovsky, James D., Andrea B. Staiti. 2005. Access and quality: does rural America lag behind? *Health Affairs* **24**(4) 1128–1139.

Ricketts, Thomas C. 1999. *Rural health in the United States*. Oxford University Press New York.

Ricketts, Thomas C. 2000. The changing nature of rural health care. *Annual review of public health* **21**(1) 639–657.

Roh, Chul-Young, Keon-Hyung Lee, Myron D. Fottler. 2008. Determinants of hospital choice of rural hospital patients: the impact of networks, service scopes, and market competition. *Journal of Medical Systems* **32**(4) 343–353.

Rosenthal, Thomas C., Chester Fox. 2000. Access to health care for the rural elderly. *JAMA* **284**(16) 2034–2036.

Rothschild, Jeffrey M., Christopher P. Landrigan, John W. Cronin, Rainu Kaushal, Steven W. Lockley, Elisabeth Burdick, Peter H. Stone, Craig M. Lilly, Joel T. Katz, Charles A. Czeisler, David W. Bates. 2005. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care*. *Critical Care Medicine* **33**(8).

Russ, Stephan, Ian Jones, Dominik Aronsky, Robert S. Dittus, Corey M. Slovis. 2010. Placing Physician Orders at Triage: The Effect on Length of Stay. *Annals of Emergency Medicine* **56**(1) 27–33.

Saghafian, Soroush, Wallace J. Hopp, Mark P. Van Oyen, Jeffrey S. Desmond, Steven L. Kronick. 2012. Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research* **60**(5) 1080–1097.

Saunders, Charles E, Paul K Makens, Larry J Leblanc. 1989. Modeling emergency department operations using advanced computer simulation systems. *Annals of Emergency Medicine* **18**(2) 134–140.

Schull, Michael J., Alex Kiss, John-Paul Szalai. 2007. The Effect of Low-Complexity Patients on Emergency Department Waiting Times. *Annals of Emergency Medicine* **49**(3) 257–264.e1.

Schull, Michael J., John-Paul Szalai, Brian Schwartz, Donald A. Redelmeier. 2001. Emergency Department Overcrowding Following Systematic Hospital Restructuring Trends at Twenty Hospitals over Ten Years. *Academic Emergency Medicine* **8**(11) 1037–1043.

Schull, Michael J., Marian Vermeulen, Graham Slaughter, Laurie Morrison, Paul Daly. 2004. Emergency department crowding and thrombolysis delays in acute myocardial infarction. *Annals of Emergency Medicine* **44**(6) 577–585.

Shapley, Lloyd S. 1952. A Value for n-Person Games. Product Page, RAND Corporation.

Shechter, Steven M., Matthew D. Bailey, Andrew J. Schaefer, Mark S. Roberts. 2008. The Optimal Time to Initiate HIV Therapy Under Ordered Health States. *Operations Research* **56**(1) 20–33.

Shi, Pengyi, Mabel C. Chou, J. G. Dai, Ding Ding, Joe Sim. 2015. Models and Insights for Hospital Inpatient Operations: Time-Dependent ED Boarding Time. *Management Science* .

Sinreich, D., Y. Marmor. 2005. Emergency department operations: The basis for developing a simulation tool. *Iie Transactions* **37**(3) 233–245.

Sinreich, David, Ola Jabali. 2007. Staggered work shifts: a way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health Care Management Science* **10**(3) 293–308.

Sinreich, David, Ola Jabali, Nico P. Dellaert. 2012. Reducing emergency department waiting times by adjusting work shifts considering patient visits to multiple care providers. *Iie Transactions* **44**(3) 163–180.

Sivey, Peter. 2012. The effect of waiting time and distance on hospital choice for English cataract patients. *Health Economics* **21**(4) 444–456.

Skandari, M. Reza, Steven M. Shechter, Nadia Zalunardo. 2015. Optimal Vascular Access Choice for Patients on Hemodialysis. *Manufacturing & Service Operations Management* .

Solberg, Leif I., Brent R. Asplin, Robin M. Weinick, David J. Magid. 2003. Emergency department crowding: Consensus development of potential measures. *Annals of Emergency Medicine* **42**(6) 824–834.

Somoza, Eugene, John R. Somoza. 1993. A Neural-network Approach to Predicting Admission Decisions in a Psychiatric Emergency Room. *Medical Decision Making* **13**(4) 273–280.

Soremekun, Olan A., Christian Terwiesch, Jesse M. Pines. 2011. Emergency Medicine: An Operations Management View. *Academic Emergency Medicine* **18**(12) 1262–1268.

Stephens, J. Mark, Samuel Brotherton, Stephan C. Dunning, Larry C. Emerson, David T. Gilbertson, David J. Harrison, John J. Kochevar, Ann C. McClellan, William M. McClellan, Shaowei Wan, others. 2013. Geographic disparities in patient travel for dialysis in the United States. *The Journal of Rural Health* **29**(4) 339–348.

Storrow, Alan B., Chuan Zhou, Gary Gaddis, Jin H. Han, Karen Miller, David Klubert, Andy Laidig, Dominik Aronsky. 2008. Decreasing Lab Turnaround Time Improves Emergency Department Throughput and Decreases Emergency Medical Services Diversion: A Simulation Model. *Academic Emergency Medicine* **15**(11) 1130–1135.

Tai, Wan-Tzu Connie, Frank W. Porell, E. Kathleen Adams. 2004. Hospital choice of rural Medicare beneficiaries: patient, hospital attributes, and the patientphysician relationship. *Health services research* **39**(6p1) 1903–1922.

Tenckhoff, H., H. Schechter. 1968. A bacteriologically safe peritoneal access device. *Transactions-American Society for Artificial Internal Organs* **14** 181.

Thompson, Matthew J., Amy Hagopian, Meredith Fordyce, L. Gary Hart. 2009. Do international medical graduates (IMGs)fill the gap in rural primary care in the United States? A national study. *The Journal of Rural Health* **25**(2) 124–134.

Tierney, William M., John Fitzgerald, Ross McHenry, Bruce J. Roth, Bruce Psaty, David L. Stump, F. Kim Anderson. 1986. Physicians' Estimates of the Probability of Myocardial Infarction in Emergency Room Patients with chest Pain. *Medical Decision Making* **6**(1) 12–17.

Tonelli, Marcello, Scott Klarenbach, Braden Manns, Bruce Culleton, Brenda Hemmelgarn, Stefania Bertazzon, Natasha Wiebe, John S. Gill, others. 2006. Residence location and likelihood of kidney transplantation. *Canadian Medical Association Journal* **175**(5) 478–482.

Tonelli, Marcello, Braden Manns, Bruce Culleton, Scott Klarenbach, Brenda Hemmelgarn, Natasha Wiebe, John S. Gill, others. 2007. Association between proximity to the attending nephrologist and mortality among patients receiving hemodialysis. *Canadian Medical Association Journal* **177**(9) 1039–1044.

Trzeciak, S., E. P. Rivers. 2003. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emergency Medicine Journal* **20**(5) 402–405.

USRDS. 2015. 2015 USRDS annual data report: Epidemiology of kidney disease in the United States. Tech. rep., National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD.

Vassilacopoulos, G. 1985. Allocating Doctors to Shifts in an Accident and Emergency Department. *J Oper Res Soc* **36**(6) 517–523.

Verter, Vedat, Sophie D. Lapierre. 2002. Location of Preventive Health Care Facilities. *Annals of Operations Research* **110**(1-4) 123–132.

Weiss, Steven J, Robert Derlet, Jeanine Arndahl, Amy A Ernst, John Richards, Madonna Fernndez-Frackelton, Robert Schwab, Thomas O Stair, Peter Vicellio, David Levy, Mark Brautigan, Ashira Johnson, Todd G Nick. 2004. Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS). *Academic emergency medicine* **11**(1) 38–50.

Wiler, Jennifer L., Richard T. Griffey, Tava Olsen. 2011. Review of Modeling Approaches for Emergency Department Patient Flow and Crowding Research. *Academic Emergency Medicine* **18**(12) 1371–1379.

Wilk, S., R. Slowinski, W. Michalowski, S. Greco. 2005. Supporting triage of children with abdominal pain in the emergency room. *European Journal of Operational Research* **160**(3) 696–709.

Wong, Hannah J, Robert C Wu, Michael Caesar, Howard Abrams, Dante Morra. 2010. Smoothing inpatient discharges decreases emergency department congestion: a system dynamics simulation model. *Emergency Medicine Journal* **27**(8) 593–598.

Woolvett, Amy. 2015a. Barrington dialysis group pressing ahead after meeting health minister. *Nova News Now* .

Woolvett, Amy. 2015b. Barrington group collecting support for dialysis clinic. *The Coastguard* .

Yankovic, Natalia, Linda V. Green. 2011. Identifying Good Nursing Levels: A Queuing Approach. *Operations Research* **59**(4) 942–955.

Yoon, Philip. 2003. Emergency department fast-track system. *Edmonton: Alberta Heritage Foundation for Medical Research* .

Zeltyn, Sergey, Yariv N. Marmor, Avishai Mandelbaum, Boaz Carmeli, Ohad Greenshpan, Yossi Mesika, Sergev Wasserkrug, Pnina Vortman, Avraham Shtub, Tirza Lauterman, Dagan Schwartz, Kobi Moskovitch, Sara Tzafrir, Fuad Basis. 2011. Simulation-Based Models of Emergency Departments: Operational, Tactical, and Strategic Staffing. *Acm Transactions on Modeling and Computer Simulation* **21**(4).

Zenios, Stefanos A., Glenn M. Chertow, Lawrence M. Wein. 2000. Dynamic Allocation of Kidneys to Candidates on the Transplant Waiting List. *Operations Research* **48**(4) 549–569.

Zhang, Yue, Oded Berman, Patrice Marcotte, Vedat Verter. 2010. A bilevel model for preventive healthcare facility network design with congestion. *IIE Transactions* **42**(12) 865–880.

Zhang, Yue, Oded Berman, Vedat Verter. 2009. Incorporating congestion in preventive healthcare facility network design. *European Journal of Operational Research* **198**(3) 922–935.

Zhang, Yue, Oded Berman, Vedat Verter. 2011. The impact of client choice on preventive healthcare facility network design. *OR Spectrum* **34**(2) 349–370.