THREE ESSAYS ON CONTINUITY OF CARE IN CANADA: FROM PREDICTIONS TO DECISIONS

THREE ESSAYS ON CONTINUITY OF CARE IN CANADA: FROM PREDICTIONS TO DECISIONS

By SOMAYE GHAZALBASH, B.Sc., M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements

for the degree of Doctor of Philosophy

Descriptive Note

McMaster University DOCTOR OF PHILOSOPHY (2022) Hamilton, Ontario

Hamilton, Ontario (DeGroote School of Business, Department of Health Policy and Management)

TITLE: Three essays on continuity of care in Canada: from predictions to decisions

AUTHOR: Somayeh Ghazalbash, B.Sc., M.Sc.

SUPERVISOR: Dr. Manaf Zargoush

NUMBER OF PAGES: 211

Lay Abstract

The aging population in Canada is growing significantly relative to the population as a whole, and several challenges are involved in providing aging people with proper healthcare services. One of these challenges is disruptions in continuity of care. Older adults are often medically complex or frail; they may have multiple diseases and require many care transitions across healthcare settings. Poor continuity of care among these patients leads to health deterioration during care trajectories, resulting in reduced quality of care and increased healthcare costs and inefficiencies. This thesis includes three essays that provide practical insights and solutions regarding the issue of continuity of care disruptions, spanning from predicting the issue to strategies to prevent it in a data-driven manner.

Abstract

Continuity of care (COC) refers to the delivery of seamless services, continuous caring relationships, and information sharing across care providers. A disruption in COC—that is, care fragmentation (CF)—is an important cause of inefficiency in the Canadian healthcare system; such disruption leads to increased healthcare costs and reduced quality of care. Addressing this issue is particularly challenging among older adults, who often have medically complex needs; such patients can require many care transitions across multiple care settings. An effective strategy for COC improvements is to optimize discharge planning among older adults. However, this is hampered by the imperfect understanding of older patients' needs, which are associated with their health complexity. Therefore, making early predictions about the patients' health complexity and incorporating this information into discharge planning decisions can potentially improve COC. In this thesis, I develop data-driven predictive–prescriptive analytics frameworks that leverage machine learning (ML) approaches and a rich, massive set of longitudinal data collected over a decade. The first essay in this dissertation studies the early prediction of older patients' complexity in hospital pathways using ML. It also examines whether we can conduct accurate prognostics with current information on patient complexity. The second study examines how two common measures of patient complexity—multimorbidity and frailty—concurrently affect post-discharge readmission and mortality among older patients. It also investigates the dependency of the outcomes on other essential socio-demographic factors. Finally, the third study examines the feasibility of predicting patients at risk of fragmented readmission—that is, readmission to a different hospital than the initial one. It uses this predictive information to derive optimal policies for preventing CF while addressing disparities in the decision-making process. The findings highlight the feasibility, utility, and performance of predicting patient complexity and important adverse outcomes, potentially undermining COC. This thesis shows that advanced knowledge and

explicit utilization of this information could support decision-making and resource planning toward a targeted allocation at the system level; moreover, it informs actions that affect patient-centered care transition at the service level to optimize patient outcomes and facilitate upstream discharge processes, thereby improving COC.

Acknowledgements

My doctoral thesis is the culmination of a long-cherished dream. There are many whose contributions I wish to celebrate at the zenith of fulfilling this dream. I extend my deepest gratitude to Dr. Manaf Zargoush, my Ph.D. supervisor, for his knowledge and expertise and for being my biggest supporter in academia and outside of it. His valuable guidance, constructive feedback, and constant support at every stage of my research ensured that I always had access to an environment rich in opportunities for learning, experimenting, teaching, reviewing, and research. Even in the most difficult moments, he helped me believe in myself, dared me to think beyond set boundaries, helped me rediscover my curious nature, and taught me to find beauty in the most unexpected of places. He's one in a million, and I am thankful for it.

I could not have undertaken this journey without the help of my committee members, Dr. Christopher Longo and Dr. Ali Reza Montazemi. The academic rigour and open conversations in our meetings have been seminal in keeping my project on track and aiming for excellence. I cannot thank them enough for being so generous with their knowledge and expertise over the course of my Ph.D. journey.

I am deeply indebted to Dr. Glen E Randall, Area Chair of the Health Policy and Management program, for his considerable guidance. I am grateful that he placed his trust in me when he accepted me as a Ph.D. candidate for the department. I am thankful for his strong support throughout.

I wish to extend my gratitude to Professors Dr. Gillian Mulvale, Dr. Neil Bar, and Dr. Jenna Evans for all the beautiful memories we created together. It has been a tough journey, but the peaceful moments that they carved out for me made it a worthwhile one. I thank them for welcoming me into their midst and making me feel at home.

I am eternally grateful to my family for being ardent supporters of this passionate endeavour. I am always in awe of and learning from my mother, Parvaneh Akbari. I will be a student of her immense patience and strength forever. My father, Kazem Ghazelbash's encouragement, has meant the world to me. My siblings' faith in me kept my spirits and motivation high. More than anything, I am thankful for the kindness they have shown me during this incredible journey.

I wish to thank the ICES support team, especially Refik Saskin and Keng-Yuan Liu, for their invaluable assistance with data and the ICES platform. The staff at DeGroote, Deb Randall, Banafsheh Rafe, Nicole Jessome, and Lindsay Ryan, have my heartfelt gratitude for their help and support. I owe a special thanks to Noora Kamar for being an amazing friend through my academic journey and in life.

I thank God for everything.

Table of Contents

List of Illustrations, Charts, Diagrams

Chapter 1 Introduction

Chapter 2 Examining the Predictability and Prognostication of Multimorbidity Among Older Delayed-Discharge Patients: A Machine Learning Analytics

Chapter 3 Impact of multimorbidity and frailty on adverse outcomes among older delayed discharge patients: implications for healthcare policy

Chapter 4 A Data-Driven Approach to Address Care Fragmentation Among Older Adults: Prediction, Decision-Making, and Fairness Considerations

List of Tables

Chapter 1 Introduction

Chapter 2 Examining the Predictability and Prognostication of Multimorbidity Among Older Delayed-Discharge Patients: A Machine Learning Analytics

Chapter 4 A Data-Driven Approach to Address Care Fragmentation Among Older Adults: Prediction, Decision-Making, and Fairness Considerations

Chapter 5 Conclusion

List of Abbreviations and Symbols

Declaration of Academic Achievement

Most of the research during my Ph.D. study (from September 2018 to August 2022) has been published or prepared for submission to peer-reviewed journals, listed as follows:

Chapter 2:

Ghazalbash, S., Zargoush, M., Mowbray, F., & Papaioannou, A. (2021). "Examining the Predictability and Prognostication of Multimorbidity Among Older Delayed-Discharge Patients: A Machine Learning Analytics." *International Journal of Medical Informatics*, 104597. DOI: https://doi.org/10.1016/j.ijmedinf.2021.104597

Chapter 3:

Ghazalbash, S., Zargoush, M., Mowbray, F., Costa, A. (2022). "Impact of multimorbidity and frailty on adverse outcomes among older delayed discharge patients: implications for healthcare policy." *Health Policy***.** DOI: https://doi.org/10.1016/j.healthpol.2022.01.004

Chapter 4:

Ghazalbash, S., Zargoush, M., Verter, V., Guilcher, S., Kuluski, K., (2022). "A Data-Driven Approach to Address Care Fragmentation Among Older Adults: Prediction, Decision-Making, and Fairness Considerations." Being prepared for submission to "*Production and Operation Management*."

Chapter 1
Introduction

Background and Motivation

Continuity of care (COC) is a multidimensional concept (Lotz 2019), including "seamless service" and "continuous caring relationship" (Gulliford et al. 2006). Seamless care refers to a smooth, integrated, and safe transition of a patient from institutional care settings (e.g., hospital) to the community (e.g., home or long-term care) (Spehar et al. 2005). Moreover, continuous caring relationship relates to the ongoing relationship between patients and care providers (individual and care facility). A temporal care provider's interaction with the same patient in the same place can create knowledge accumulation about the patient across care episodes, develop a trusting bond between them, and enhance information sharing, hence improving informational continuity (i.e., sharing information across the providers) (Senot 2019).

A disruption in the COC and compromised care coordination among providers, referred to as care fragmentation (CF), is one of the main causes of the healthcare system's inefficiency, leading to increased healthcare costs (Rosenberg and Zulman 2020) and reduced quality of care (Hirji et al. 2020). It also contributes to the progression of patient comorbidities and their dissatisfaction (Crilly et al. 2006; Hirji et al. 2020; Juo et al. 2019; Rosenberg and Zulman 2020). CF becomes more serious among older patients, who are often medically-complex with multiple chronic conditions and geriatric syndromes, leading to an increased need to be seen by various care providers, transfer across multiple care settings, and efficient core coordination (Brooke 2020).

An effective strategy for improving CF, particularly among the medically-complex aging population, is enhancing transitional care programs and optimizing discharge planning (Lee et al. 2022; Rasmussen et al. 2021). The transition of care, i.e., moving patients between various levels of health care and across different care settings (Naylor and Keating 2008), is challenging and

costly (Nasarwanji et al. 2015). Suboptimal care transitions increase the risk of post-discharge adverse events, such as complications, mortality, and rehospitalization (Hoyer et al. 2018). As such, timely, proactive, and informed discharge planning can facilitate a safe and smooth care transition (Fox et al. 2013; Ohta et al. 2016; Spehar et al. 2005). One barrier to this optimal planning is an imperfect understanding of patients' distinct physical and psychosocial needs associated with their complexity (Bakerjian et al. 2019). Therefore, early predictions of the patients' health complexity in the hospital pathways and incorporating this information into the discharge planning practices can significantly contribute to managing CF among older patients. However, despite the availability of extensive data (e.g., electronic medical records) and the exponential surge in computational power and data storage over the past decade, few efforts have been made to predict patient complexity, particularly using machine learning (ML).

After capturing and predicting patient complexity (i.e., ex-ante), it becomes critical to conduct accurate prognostications with this information (i.e., ex-post). Here, prognostication refers to the strength of the risk factors in predicting patient outcomes (Lønning 2007). Advanced knowledge of the effects of patient complexity on adverse outcomes among older patients could help stratify patients at risk for adverse health events and support proactive discharge planning (and other interventions) both in the hospital and following discharge, supporting care continuity. It can also enhance shared decision-making. A US study indicated that up to 70% of patients might change their decisions and preferences about their care plan after they have been fully informed about their probable outcomes and were engaged more in the process of decision-making about their care (Robinson and Jagsi 2016).

Although readmission can negatively affect COC, disjointed/disrupted care because of discontinuity between care facilities—that is, fragmented readmission—can potentially exacerbate CF and worsen outcomes as a result (Juo et al. 2019). For example, one study reported that the average cost difference between fragmented and non-fragmented readmissions ranges from $270 to $22,000 per patient (Snow et al. 2020). Hospitals are more likely to have common physical, electronic, and managerial systems to facilitate the sharing of critical patient information, such as admission, medical history, complications, progress notes, diagnostic test results, and in-hospital therapies or interventions. Moreover, receiving care from the same hospital increases team familiarity and informal information sharing across different providers, leading to improved informational continuity (Senot 2019). One strategy to reduce fragmented readmissions is to predict this event among patients and impose targeted interventions accordingly (Agha et al. 2019; Hirji et al. 2020). Most existing research focusing on such approaches has applied statistical methods (in a relatively highly theoretical framework) for association analysis to compare index and non-index readmissions (Ando et al. 2019). They have also measured the concentration level of patient care episodes in the same care setting to investigate its association with outcomes. However, given the exponential advances in data collection, data storage, and computational power over the past decade, the existing literature has not thoroughly identified a predictive approach to provide actionable insights into the issue of fragmented readmission.

My first essay in this dissertation facilitates early predictions of the patients' health complexity in the hospital pathways (Ghazalbash et al. 2021). Patient complexity among older patients

complicates discharge planning, resulting in a higher rate of adverse outcomes, such as readmission and mortality. As a common indicator of patient complexity, early prediction of multimorbidity can support proactive triage and decision-making about staffing and resource allocation to optimize patient outcomes. Moreover, it facilitates an upstream and informed discharge process by prioritizing complex patients for discharge and providing patient-centered care; this improves seamless care, which is a fundamental component of COC.

In the first essay (i.e., Chapter 2), we examine the predictability of three common multimorbidity indices, including the Charlson–Deyo Comorbidity Index (CDCI), the Elixhauser Comorbidity Index (ECI), and the Functional Comorbidity Index (FCI) using ML. Moreover, we assess the prognostic value of these indices in predicting 30-day readmission and mortality. In particular, we focus on delayed-discharge patients, referred to in Canada as alternate level of care (ALC) patients. While awaiting enrollment in community-based services, ALC patients occupy an acute or post-acute bed despite no longer requiring the intensity of these resources. They receive their needed care with a delay until they are transferred to the appropriate alternate facility. This delay undermines the "seamless care" component of COC. To our knowledge, this is the first study to thoroughly and concurrently investigate the ML predictability of patient complexity via multimorbidity indices and their prognostic value for predicting patient-important outcomes among older adults or those facing delayed hospital discharge. We provide first-hand evidence on this topic through robust comparisons of several predictive accuracy measures using multiple ML algorithms with three common multimorbidity indices. Our results indicate that regardless of the

type of the ML algorithm or the measure for predictive performance assessment, the CDCI is the most predictable index, whereas the FCI is the least predictable. Interestingly, these two indices had the least agreement regarding their predictions. More remarkably, the prognostication analytics in our study revealed that the most predictable index (i.e., the CDCI) also had the greatest strength in predicting adverse events.

The differences between multimorbidity indices (in terms of both predictability and prognostication), as well as the degree of their prediction agreements, may stem from the type of comorbidity items included in the index and the way the items are combined. Although the CDCI consists of fewer comorbidities compared with the ECI and FCI, the combination of the comorbidity items in the CDCI is weighted based on their severity to account for the disease burden, which contrasts with the simple summation of the comorbidity items in the ECI and FCI. We also contribute to this literature by highlighting the importance of accounting for the severity of diseases for better prognostications with multimorbidity indices; we are the first to highlight such importance for improving the predictability of multimorbidity indices.

My second essay examines how two measures of patient complexity—that is, multimorbidity and frailty—concurrently affect the 30-day post-discharge readmission and mortality among older patients and their dependency on other important factors (Ghazalbash et al. 2022). This leads to insights into evidence-based policies for improving patient care transitions. Through detailed characterization of the relationship between concurrent multimorbidity and frailty with patient-important outcomes, the paper provides policy-related and managerial insights to understand better

the complex needs of older patients, particularly those with delayed discharge. It also informs discharge policies accordingly.

Although several studies have concluded that multimorbidity is the most commonly used component of patient complexity, as discussed in my first essay, such geriatric syndromes as frailty are considered yet another critical dimension of patient complexity in the context of the aging population. Frail patients with multiple chronic conditions (multimorbidity) are at greater risk for adverse outcomes, such as readmission; this increases the risk of disrupted care continuity. This study focused on delayed-discharge patients because the co-occurrence, severity, and consequences of frailty and multimorbidity are likely greater among delayed-discharge patients, highlighting a cohort of older persons with mostly complex conditions following an acute illness. Both in the hospital and following discharge, advanced knowledge of the effects of multimorbidity and frailty on adverse outcomes among these patients could help stratify patients at risk for adverse health events and support proactive discharge planning and other interventions.

The aim of the second essay (i.e., Chapter 3) is to examine the coexisting effects of multimorbidity and frailty measured through the electronic health record (EHR)-based ECI and the Hospital Frailty Risk Score (HFRS), respectively, on a series of patient outcomes through an extensive analysis of large data from a Canadian population. Further, the study aims to provide insights for enhancing discharge planning and resource allocation toward a smoother care stream. Our findings indicate that multimorbidity and frailty provide unique information about adverse outcomes among older patients with delayed discharge, but they are most informative when examined in tandem.

To complete this picture, we also examine the dependency of the outcomes on other important factors, such as sex, rural/urban residency, marginalization status, and wait time. These investigations lead to the critical health policy implications discussed in the paper. Advanced knowledge of these factors could support proactive, evidence-informed, and equitable discharge planning and clinical decision-making regarding discharge delay.

Finally, the third essay examines the feasibility of predicting care fragmentation (CF) and investigates how prediction models can be used for targeted interventions in clinical practice via extensive comparison with random intervention strategies. In this essay, presented in Chapter 4, we develop a data-driven predictive–prescriptive analytics framework that leverages big data and ML techniques to predict patients at risk of fragmented readmission—that is, readmission to a different hospital from the initial hospital, a key aspect of CF. Moreover, we derive optimal intervention policies for preventing CF while addressing disparities in decision-making. We utilize a rich set of longitudinal data collected over a decade, with approximately 1 million unique observations, to develop ML-based predictive models for fragmented readmission. Our data also record the cost of hospitalization at the patient visit level, allowing us to estimate personalized data-driven hospital expenses. Our predictive models seek to mitigate two challenges of previously published ML-based predictive models, which are as follows: a) discriminative decisions pertaining to underserved patients stemming from algorithmic biases (Fu et al. 2020) and b) a lack of cost-based metrics to evaluate the clinical impact of predictive models in healthcare decision-making (Teo et al. 2021). Although each challenge has been investigated separately in the current

ML-based healthcare decision-making literature, this study aims to overcome both challenges simultaneously to provide a more robust and practical decision-making framework. Finally, we explicitly link our proposed ML predictive analytics with prescriptive analytics to formulate a data-driven optimization framework. This framework renders the optimal range of intervention costs through which an ML-based strategy is beneficial in terms of cost savings regarding the strategies in current clinical practices. From a more practical standpoint, we provide a clinical decision-making framework that assists both service- and system-level decision-makers. At the service level, this suggests an optimal screening policy for targeting patients with a high risk of CF while balancing the tradeoff between the cost of wrong screening (because of the ML's false predictions) and potential cost savings regarding the current random strategies. At the system level, it assists policymakers in selecting optimal preventive interventions, given the available budget, intervention costs, and effectiveness, for varying degrees of risk aversion. This study provided first-hand evidence of accurate fragmented readmission prediction via robust comparisons of multiple predictive models using different performance measures. Further, our results indicated that the effectiveness of the ML-based strategy would depend on the performance of risk models in predicting positive and negative cases, the effectiveness of interventions, the cost of the readmission (fragmented and non-fragmented), the rate of fragmented readmission among subgroups, and the degree of risk aversion. We showed that utilizing our predictive–prescriptive framework can potentially yield significant financial savings. It can also facilitate shared decision-making in inpatient care settings by providing a platform to communicate information about the risk of CF and discharge planning to patients, their families, and care providers. In this way,

patients and their families can be informed of the risk of CF and decide on more engagement in the process of following up after discharge, leading to decreased fragmentation of care. Finally, our framework accounts for fairness considerations, which can have far-reaching policy implications by allowing the disparity learned from data to be computed and by correcting unfair decisions toward a fair ML-based decision.

Research Streams and Content Overlap in the Thesis

This thesis is relevant to the following research fields: a) aging research, b) patient complexity, c) transitional care delivery (particularly COC) on the context side, and d) data-driven predictive–prescriptive analytics on the methodological side. Figure 1 depicts the positions of the three studies in the relevant literature, and the contents overlap across the three studies. The first study, leveraging ML-based predictive analytics, contributes to the aging, patient complexity, and COC literature. The study evidenced the ML predictability of patient complexity based on multimorbidity indices and their prognostic significance in predicting patient-important outcomes, such as mortality and readmission among older adults or those facing delayed hospital discharge. It provides first-hand evidence on this topic via robust comparisons of several predictive accuracy measures using multiple ML algorithms with three common multimorbidity indices. The second study contributes to the aging, patient complexity, and COC literature (particularly seamless services), leveraging statistics-based predictive analytics. In this study, through an extensive analysis of big data from a Canadian population, we examined the coexisting effects of multimorbidity and frailty, as two components of patient complexity, on a series of patient

outcomes. This is the first study to examine the effects of multimorbidity and frailty on a series of patient-important outcomes, including mortality and hospital readmission, among older adults with delayed discharge. The third study contributes to aging, patient complexity, and COC (particularly fragmented readmission), leveraging ML-based predictive-prescriptive analytics. We developed a data-driven predictive–prescriptive analytics framework that leverages big data and ML methodologies to drive optimal intervention policies for preventing CF while addressing disparities in the decision-making process. In this study, we developed a competitive ML-based prediction model to identify patients at risk of fragmented readmission. Furthermore, we illustrated and investigated how ML predictions can be used for targeted interventions in real clinical practices through extensive comparisons with random intervention strategies. Finally, we examined the fairness implications of the developed ML-based decision-making framework to ensure parity among protected groups.

To our knowledge, no extant studies have addressed the problem of predicting CF among the older population, and prior studies have not investigated the value of optimal ML-based strategies for preventive interventions compared with existing clinical practices. In addition, unlike the current studies that measure the concentration level of patient care episodes in the same physical location to investigate its association with outcomes, our study aimed to predict the fragmentation of continued care in terms of physical location—that is, readmission to any hospital other than the hospital at which the initial care was received. We capitalized on these predictions to design optimal preventive strategies with the potential to save healthcare costs.

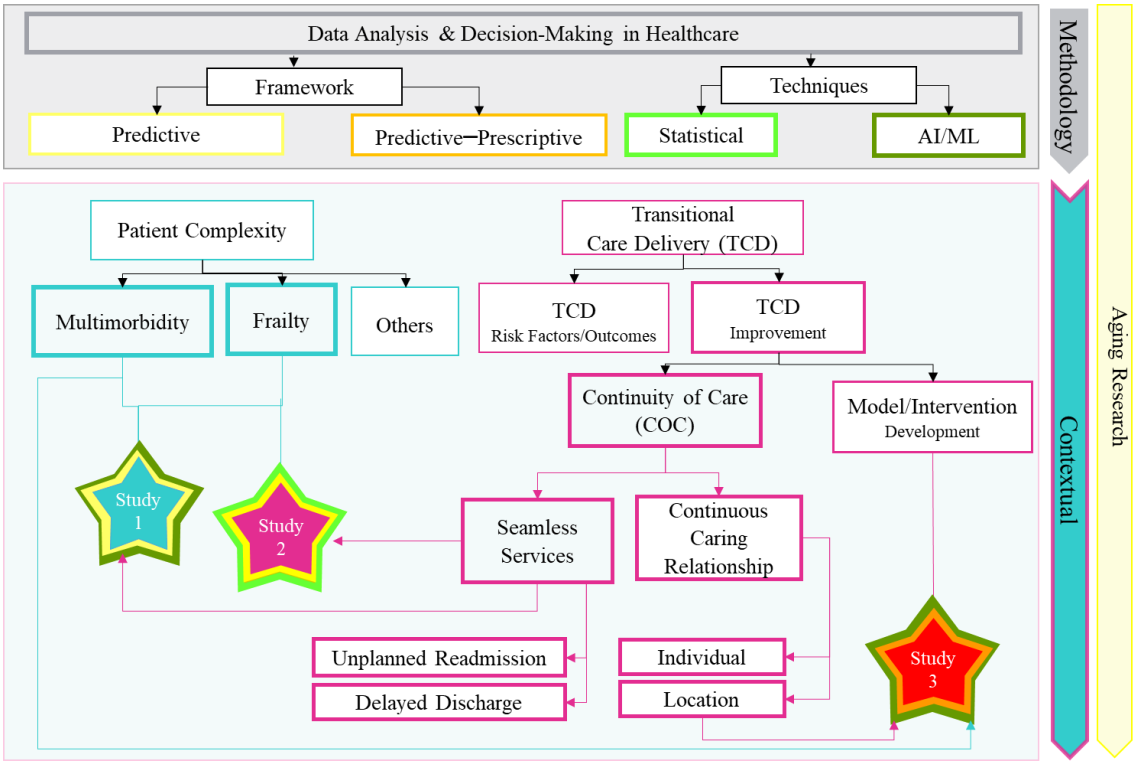


Figure 1 Research streams and content overlap across studies

Figure 2 also illustrates the relationship between the contents of the above-mentioned streams in this thesis. As the figure shows, there is a bidirectional relationship between patient complexity and transitional care delivery. Advanced knowledge of patient complexity can improve the transition of care programs and discharge planning. Consequently, it can potentially reduce the risk of readmission, particularly among delayed-discharge patients (as discussed in the first and second essays), which can also lead to CF (the third essay). In contrast, a suboptimal transition of care/discharge planning can lead to health deterioration in terms of multimorbidity and functional impairment (i.e., patient complexity; Van Cleave et al. 2013). I used the data and decision analytics

approaches to find solutions that tackle the abovementioned issues and provide managerial and policy insights.

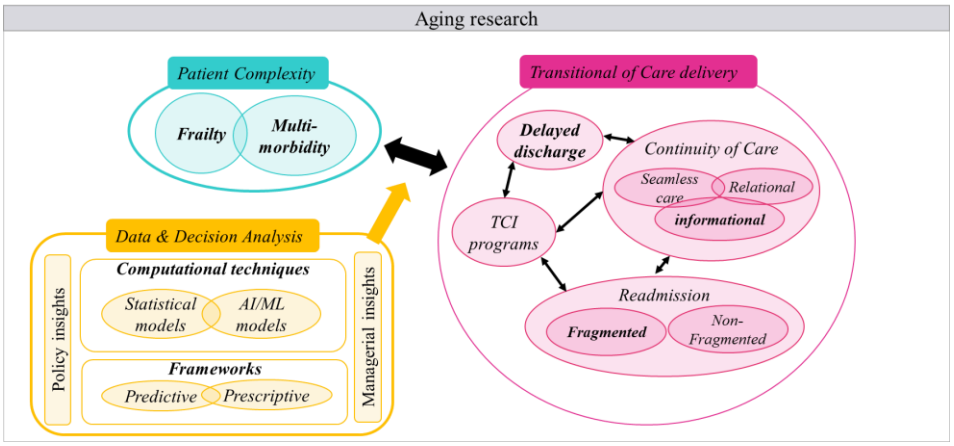


Figure 2 Relationship between the contents of research streams

Data Availability and Potential Sources

Three retrospective cohort studies were conducted using data extracted from several health administrative databases. These databases are provided by the Canadian Institute for Health Information (CIHI) and housed at the Institute for Clinical Evaluative Sciences (IC/ES). All residents aged 65 years and older who experienced an episode of care between 2004 and 2017 in Ontario, Canada, were eligible for study inclusion. This study was approved by the Hamilton Integrated Research Ethics Board (HiREB).

In this thesis, the following databases are used: the Discharge Abstract Database (DAD), the National Rehabilitation Reporting System (NRS), and a cohort derived from the Registered Persons Database (RPDB). In the study, the databases were linked at the episode level using a unique encrypted identifier that allowed longitudinal tracking of patient care journeys across

multiple settings. Table 1 presents more details regarding the databases**.** Figure 3 illustrates the possible transitions across different care facilities and related IC/ES databases. In three studies, I used the databases as follows: a) DAD to capture hospitalization discharge information; b) the NRS to extract data on admission and discharge from the inpatient rehabilitation program; and c) the RPDB to capture demographic and socioeconomic variables, such as age, sex, rurality of residence, and the neighbourhood marginalization index.

Table 1 Detailed information on IC/ES datasets aimed to use in this study

Database	Main information	# Records	#Variables (Clinical & Administrative)
DAD	Patient visits records to acute and chronic care hospitals	11,014,847	445
NRS	Patient visits records to inpatient rehabilitation centers	339,117	113
RPDB[§]	Patient's demographic information	5,042,000	54

[§] We have access to a selected cohort of this dataset

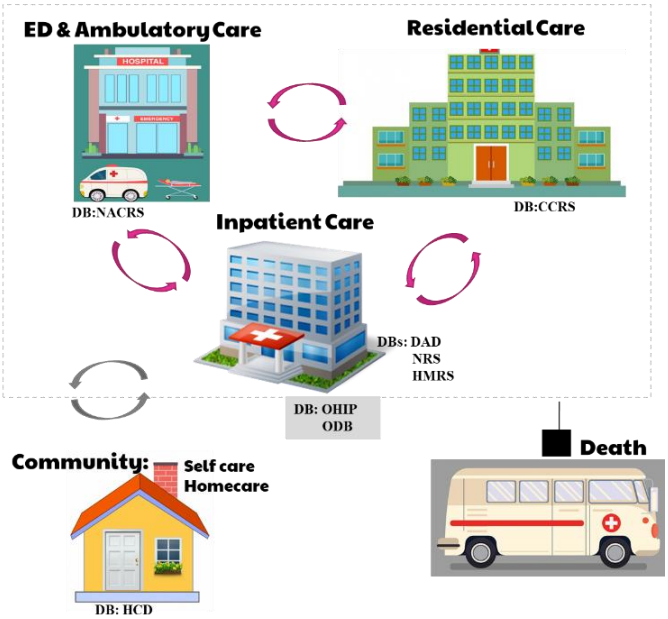


Figure 3 A scheme of transitional care: the passage between levels of health care and across care settings

Note: DB refers to the database; Continuing Care Reporting System (CCRS); Home Care Database (HCD); Ontario Health Insurance Plan Claims Database (OHIP); Ontario Drug Benefit Claims (ODB); Ontario Mental Health Reporting System (OMHR)

Thesis Outline

The remainder of the thesis is organized as follows. Chapter 2 presents the first essay: "Examining the Predictability and Prognostication of Multimorbidity Among Older Delayed-Discharge Patients: A Machine Learning Analytics." Chapter 3 presents the second essay: "Impact of multimorbidity and frailty on adverse outcomes among older delayed discharge patients: implications for healthcare policy." Chapter 4 presents the third essay: "A Data-Driven Approach to Address Care Fragmentation Among Older Adults: Prediction, Decision-Making, and Fairness Considerations." Chapter 5, i.e., the concluding part, demonstrates how the three essays are interconnected and describes the practical implications of the results. It also provides possible research proposals for future works.

Chapter 2

Examining the Predictability and Prognostication of Multimorbidity Among Older Delayed-Discharge Patients: A Machine Learning Analytics

Ghazalbash, S., Zargoush, M., Mowbray, F., & Papaioannou, A. (2021). "Examining the Predictability and Prognostication of Multimorbidity Among Older Delayed-Discharge Patients: A Machine Learning Analytics". **International Journal of Medical Informatics**, 104597. DOI: https://doi.org/10.1016/j.ijmedinf.2021.104597

Preface

The first article in this dissertation, entitled “Examining the Predictability and Prognostication of Multimorbidity Among Older Delayed-Discharge Patients: A Machine Learning Analytics,” was published in the International Journal of Medical Informatics (IJMI) in December 2021. In this research, research design and problem formulation were conducted by S. Ghazalbash and M. Zargoush. Overseen by M. Zargoush, data analytics were conducted by S. Ghazalbash, who also drafted the initial version of the manuscript. S. Ghazalbash and M. Zargoush contributed to interpreting the results and revised the manuscript for intellectual content. A. Papaioannou and F. Mowbray provided insights regarding the clinical and policy implications of the results. M. Zargoush oversaw the entire study.

Abstract

Background: Patient complexity among older delayed-discharge patients complicates discharge planning, resulting in a higher rate of adverse outcomes, such as readmission and mortality. Early prediction of multimorbidity, as a common indicator of patient complexity, can support proactive discharge planning by prioritizing complex patients and reducing healthcare inefficiencies.

Objective: We set out to accomplish the following two objectives: 1) to examine the predictability of three common multimorbidity indices, including Charlson–Deyo Comorbidity Index (CDCI), the Elixhauser Comorbidity Index (ECI), and the Functional Comorbidity Index (FCI) using machine learning (ML), and 2) to assess the prognostic power of these indices in predicting 30-day readmission and mortality.

Materials and Methods: We used data including 163,983 observations of patients aged 65 and older who experienced discharge delay in Ontario, Canada, during 2004 – 2017. First, we utilized various classification ML algorithms, including classification and regression trees, random forests, bagging trees, extreme gradient boosting, and logistic regression, to predict the multimorbidity status based on CDCI, ECI, and FCI. Second, we used adjusted multinomial logistic regression to assess the association between multimorbidity indices and the patient-important outcomes, including 30-day mortality and readmission.

Results: For all ML algorithms and regardless of the predictive performance criteria, better predictions were established for the CDCI compared with the ECI and FCI. Remarkably, the most predictable multimorbidity index (i.e., CDCI with Area Under the Receiver Operating Characteristic Curve = 0.80, 95% CI = 0.79 – 0.81) also offered the highest prognostications regarding adverse events ($RRR_{mortality}$ = 3.44, 95% CI = 3.21 – 3.68 and $RRR_{readmission}$ = 1.36, 95% CI = 1.31 – 1.40).

Conclusions: Our findings highlight the feasibility and utility of predicting multimorbidity status using ML algorithms, resulting in the early detection of patients at risk of mortality and readmission. This can support proactive triage and decision-making about staffing and resource allocation, with the goal of optimizing patient outcomes and facilitating an upstream and informed discharge process through prioritizing complex patients for discharge and providing patient-centered care.

Keywords: *multimorbidity; machine learning prediction; prognostication; patient complexity; older adults; delayed discharge*

1. Introduction

Limited availability of community-based services, particularly long-term and continuing care, impedes the optimal discharge of older adults, resulting in their delayed hospital discharge. Delayed discharge, also known as bed-blocking, is a prevalent healthcare system issue in many countries, including Canada, the United Kingdom, the United States, Norway, New Zealand, Australia, and Sweden [1–3], where one-third of hospitalized older adults experience discharge delay [4]. In Canada, these patients are referred to as alternate level of care (ALC) patients. While awaiting enrollment in community-based services (such as home care, rehabilitation, or long-term care), they occupy an acute or post-acute bed despite no longer requiring the intensity of these resources [5]. Whereas delayed discharge's direct effects primarily relate to older patients, this issue also has a domino effect across the healthcare system, affecting all patient populations. Among older hospitalized patients, it is strongly associated with a decline in overall health, functional mobility, and adverse medical events, such as infection, falls, and delirium [6–8]. Delayed discharges, then, affect patient flow across the healthcare system, resulting in increased emergency department wait times and outpatient elective surgery cancellations [6,7]. These consequences, therefore, lead to inefficient use of acute care beds, significant increases in healthcare costs, and the system's inefficiency [3,4].

One possible solution to the delayed discharge issue and its ramifications is to increase bed capacity in community-based services. Nevertheless, this is a costly solution, which may not be feasible in the short or medium term. An alternative solution is to improve discharge policies through optimizing capacity planning and resource allocations in community-based services [9]. However, a key barrier to this solution is an imperfect understanding of the distinct physical and

psychosocial needs of the delayed-discharge patients associated with their complexity [10]. Therefore, early predictions of the patients' health complexity in the hospital pathways and incorporating this information into the discharge planning practices can significantly contribute to the management of healthcare services among older delayed-discharge patients.

Multimorbidity, defined as the coexistence of two or more diseases [11–13], is the most common measure of patient complexity [14]. It offers unique prognostic values for determining patient outcomes and making clinical decisions regarding hospitalized older patients [15]. Patients with multimorbidity are characterized as complex patients based on their extensive and heterogenous healthcare needs across various clinical specialties [16,17]. Multimorbidity also complicates self-management in the home and continuity-of-care among specialty services, adding further complexity to patient management and system planning [18,19]. Moreover, a compelling body of literature indicates that healthcare costs exponentially increase with the multimorbidity level [16,20]. Therefore, early consideration of multimorbidity as an indicator of patient complexity can lead to a better understanding of the complex needs of older delayed-discharge patients and improve their discharge by prioritizing those at greater risk of adverse outcomes.

There are various measures of multimorbidity, such as the Charlson–Deyo Comorbidity Index (CDCI) [21], the Elixhauser Comorbidity Index (ECI) [22], and the Functional Comorbidity Index (FCI) [23]. These composite health indicators are brief and informative measures to promptly determine the overall patient health and complexity, and they are strongly associated with patient vulnerability and health service use [24–26]. Despite the availability of large data (e.g., electronic medical records) and the exponential surge in computational power and storage over the past decade, few efforts have been made to predict patient multimorbidity using machine learning

(ML). Moreover, little is known about how the predictability of multimorbidity indices compares with their prognostications. To fill these gaps, we set out to accomplish a predictability examination of three common multimorbidity indices using ML and assess the prognostication of these indices in predicting 30-day readmission and mortality. We elected to examine the above indices of multimorbidity status, bearing in mind that they are easily calculated, and the diagnostic history is readily available during patient admission assessment, or if needed, in medical records. In this study, predictability means the extent of predictive performance with which the patients' multimorbidity status can be predicted. Prognostication, on the other hand, refers to the strength of the multimorbidity indices in predicting patient outcomes, i.e., 30-day readmission and mortality [27].

2. Related works

Prior studies have mainly used descriptive analytics regarding the causes and consequences of discharge delay among older adults [28–30]. Moreover, predictive analytics have been used to examine reducing patients' length of stay [31] and early identification of patients at risk of delayed discharge [5]. The current studies on predicting multimorbidity conditions primarily focus on identifying the determinants of multimorbidity, e.g., sociodemographic, health, and lifestyle factors [32–35]. These studies often consider certain illnesses and use statistical analyses of small samples. However, little is known about the utility of ML for predicting multimorbidity among older patients, particularly those experiencing delayed hospital discharge, or assessing which multimorbidity measure can best improve these predictions. In a longitudinal study, Shang and colleagues [36] employed ML algorithms, including logistic regression, random forest, gradient boosting, and deep learning, to identify the leading multimorbidity predictors. The study has

defined multimorbidity as the concurrence of multiple (≥2, ≥3, and ≥4) conditions among 11 reported comorbidities. A recent review study examined the applications of ML in exploring the various patterns of multimorbidity evolutions [14]. The literature on the prognostication of multimorbidity indices is richer than the other bodies of the research described above [37–45]. However, none of the studies with this literature have been conducted regarding discharge delay.

3. Materials and methods

3.1. Proposed methodology

Figure 1 summarizes the data analysis framework for the predictability and prognostication analytics in this study. In summary, after data preparation in Step 1, the hyperparameters were tuned to identify the best model in Step 2. Then, in Step 3, we evaluated the performance of the best model on the test set. Finally, in Step 4, we used multinomial logistic regression to assess the association between multimorbidity status and adverse outcomes. All analyses were implemented using R 3.3.1. The "caret" package was used to implement predictive models and conduct model selection, and "nnet" was applied for data preprocessing. Several other packages were used for hyperparameter tuning. Below, we provide further details about various components of the methodological framework.

3.2. Data sources

We carried out a retrospective cohort study of individuals aged 65 and older who experienced delayed hospital discharge in Ontario, Canada's largest province, between April 2004 and March 2017. Three databases, housed in the Institute for Clinical Evaluative Sciences (IC/ES), were used to create the study data. The Discharge Abstract Database (DAD) and the National Rehabilitation Reporting System (NRS) were used to extract data regarding patient health status at admission and

discharge as well as health service use. The patients' sociodemographic characteristics were extracted using the Registered Persons Database (RPDB).

Patients were eligible for inclusion in the analysis if they were labeled alternate level of care and discharged from the hospital or died while awaiting discharge, leading to 164,454 observations. After removing patients with missing data (less than 0.3% predating to the marginalization variables only), the final study sample included 163,983 observations. For the predictability analytics, we excluded patients who died during the hospitalization, resulting in 131,699 records. For the prognostication analytics, we selected patients with at least two visits to capture patient readmission, resulting in 96,443 records. Appendix A displays the data eligibility flow diagram in our study. This study was approved by the ethical board of the Hamilton Integrated Research Ethics Board (HiREB) in Ontario.

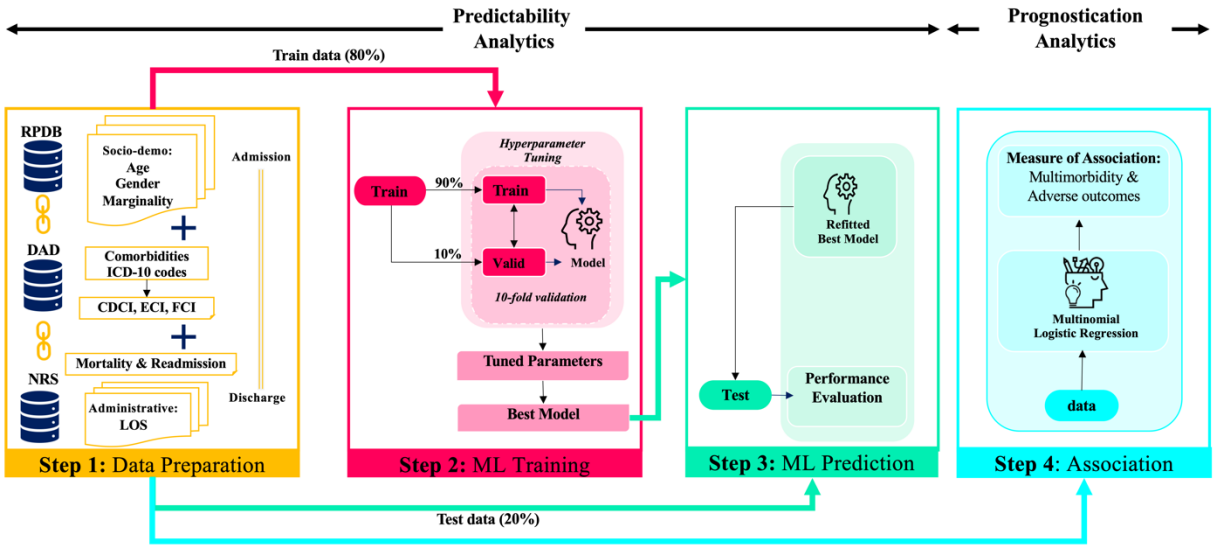


Figure 1. The methodological framework used in this study

3.3. Data descriptions

Although prediction models can be used at any point of patient contact, they provide the greatest utility when they can support real-time clinical decision-making. To maintain the feasible

implementation of the algorithms upon admission to an acute care setting, we included patient and clinical factors readily available to healthcare providers and policymakers upon hospital admission assessment or administrative intake to highlight the viability of screening. More specifically, we used data readily available to clinicians (e.g., age, sex, and diagnoses) and policymakers (e.g., readmission status and admission type) to support real-time, proactive decision-making concerning clinical management and discharge planning upon transfer from the emergency department.

Candidate predictor selection was guided by prior work and data availability. Predictor variables include the following: (i) patient demographics (age, gender), (ii) socioeconomic status (marginalization and rural/urban residency), (iii) acute and chronic conditions, and (iv) administrative variables (hospital length of stay, admission type (urgent vs. non-urgent), and readmission history indicator. Marginalization was measured using the Ontario marginalization index [46], which consists of material deprivation, residential instability, ethnic concentration, and dependency. These items are measured on a quantile scale from Q1 to Q5, where Q5 represents the most severe level of marginalization. Admission and discharge diagnoses were determined using the primary International Classification of Diseases, 10th edition (ICD-10; Canadian version). The acute and chronic conditions are limited to those included in the definitions of three common comorbidity indices, i.e., CDCI, ECI, and FCI. The details regarding the acute and chronic conditions included in the multimorbidity indices are available in Appendix B.

This study included two types of outcome variables for the two conducted analytics. The primary outcome for the predictability analytics was the multimorbidity status at 30 days post-

admission, and the secondary outcomes for the prognostication analytics were readmission and mortality within 30 days post-admission.

3.4. Data preprocessing and preparation

To measure multimorbidity status, we used the CDCI, ECI, and FCI. The first two indices are commonly used to predict all-cause mortality, length of stay, and healthcare-related costs [37]. The CDCI is a weighted sum of 17 comorbidities, whereas the ECI is a simple sum of 31 comorbid conditions. In contrast, the FCI was designed to predict physical function and is calculated as a sum of 18 comorbid conditions [47]. The details of diseases mapping with the Canadian ICD-10 codes can be found in [38]. The dichotomization threshold of two was determined to be the best cut-off point for discriminating patients on the two categories, which can be interpreted as *severe multimorbidity status* (index ≥ 2) versus *non-severe status* (index = 0 – 1) [48]. Using the binning method, we discretized the age variable into four intervals: 65–71, 71–77, 77–83, and ≥83. The readmission history was dichotomized based on whether the patient was admitted for the first time or not.

3.5. Predictive models

To assess the predictability of multimorbidity indices, we used five supervised ML algorithms: (i) classification and regression tree (CART), (ii) random forest (RF), (iii) extreme gradient boosting (XGB), (iv) bagging tree (BAG), and (v) logistic regression (LR). The tree-based algorithms have performed well in clinical and administrative practice, such as early identification of hospitalization [49], prediction of patients at the risk of stroke at emergency department triage [50], and predicting triage level [51]. They do not require statistical assumptions such as the absence of multicollinearity, they are simple and robust to outliers, and can capture the complex

non-linear relationships [49,52–54]. Some studies have also shown that LR can perform as well as ML algorithms in predicting acute kidney injury [54] or patient mortality [55]. For the predictability analytics, the dataset was randomly divided into a training sample (80%) and a test sample (20%) to train and evaluate ML algorithms. All ML algorithms were also trained using 10-fold cross-validation to determine the optimal hyperparameter values and avoid overfitting.

For the prognostication analytics, we used multinomial LR, adjusted for all covariates, with three outcomes: mortality (coded as outcome = 1), readmission (outcome = 2), or neither (outcome = 0) within 30 days after admission. We used the relative risk ratio (RRR) to measure association, using no-event (i.e., outcome = 0) as the reference outcome. For this analytics, the absence of multicollinearity was examined through the variance inflation factor (VIF) of the predictors, and the Akaike information criterion (AIC) was used as a goodness of fit measure.

3.6. Performance evaluations

The discrimination power was assessed using various measures, including (a) recall (sensitivity) to evaluate the effectiveness of the classifier to identify the positive label (i.e., severe multimorbidity), (b) precision to evaluate the ability of the classifier to avoid the wrong prediction in the positive label, (c) area under the receiver operating characteristic, i.e., ROC, curve (AUC) to evaluate the tradeoff between true-positive and false-positive rates, (d) F1-measure to combine the values of recall and precision through a harmonic means, and (e) accuracy to evaluate the overall prediction accuracy of the classifier on the test data [56,57]. Finally, to assess the calibration as well as the clinical utility of the predictive models, we conducted analyses of calibration curves and decision curves, respectively. The calibration curve assesses how well the predicted probabilities agree with the observed probabilities [58]. The potential clinical utility of

the predictive models was evaluated a range of threshold probabilities. The net benefit is equal to the expected benefit to the cases (i.e., true positive rate) minus the expected harm to the controls (i.e., false positive rate multiplied by the threshold probability) [59]. For ease of interpretation, we utilized the standard net benefit, which has a maximum value of 1.0 [60].

3.7. Hyperparameter tuning and model selection

Grid search was employed to explore the hyperparameter space for the optimal ML hyperparameter values. For each algorithm, we defined a combination profile for hyperparameter values. Then, we evaluated the algorithm's performance under each profile using the 10-fold cross-validated AUC and selected the one yielding the highest value. For CART, the complexity parameter (cp) was tuned using the "rpart" package, and for RF, the number of randomly selected predictors (mtry) was tuned using the "randomForest" package. The six following hyperparameters were tuned for XGB using the "xgboost" package: (i) number of boosting iterations (NBI), (ii) max tree depth (MTD), (iii) shrinkage/learning rate (eta), (iv) minimum loss reduction (gamma), (v) subsample ratio of columns (SRC), and (vi) minimum sum of instance weight (MSIW). We applied the grid search iteratively to shrink the range of searches and improve the results based on the previous iterations. The details of the hyperparameter tuning are available in Appendix C.

4. Results

4.1. Descriptive results

The results of summary statistics are provided in Table1. The study included a total of 163,983 observations of patients aged 65–102. The mean age (standard deviation) of the patients was 77.1 (7.9) years, and over 39% were older than 80. Most patients were female (55.4%), admitted from

an urban area (90.4%), had urgent admission with triage levels of 1–3 (85.9%), and were admitted through the emergency department (68.7%). In the extracted data, 19.7% of patients died in the hospital, and the proportion of the patients with severe multimorbidity status was 38% based on the CDCI, 46% based on the ECI, and 42% based on the FCI at the discharge time. At the admission, the CDCI scores ranged from 0 to 18 with a mean of 1.28 (1.91), the ECI scores ranged from 0 to 12 with a mean of 1.11 (1.27), and the FCI scores ranged from 0 to 8 with a mean of 0.97 (0.92). At the discharge, the CDCI scores ranged from 0 to 24 with a mean of 2.14 (2.48), the ECI scores ranged from 0 to 14 with a mean of 2.46 (2.40), and the FCI scores ranged from 0 to 11 with a mean of 1.87 (1.82). We examined the statistical significance of the differences between the index means at the admission and discharge (details are available in Appendix D). The results reveal that all the differences are statistically significant (p-value<0.0001), providing solid evidence for the increased multimorbidity burden from admission to discharge among older delayed-discharge patients.

Table 1. Descriptive details regarding the patient characteristics

Characteristics	Number (%) / Mean (Std Dev)
Age, mean (SD)	77.12 (7.9)
65-71	46,036 (28%)
71-77	39,031 (24%)
77-83	41,621 (25%)
≥83	37,295 (23%)
Sex	
Male	73,134 (44.6%)
Female	90,849 (55.4%)
Residency	
Rural	15,700 (9.6%)
Urban	148,283 (90.4%)
Material deprivation	
Quintile 1(least)	25,419 (15.5%)
Quintile 2	30,328 (18.5%)
Quintile 3	31,936 (19.5%)
Quintile 4	36,588 (22.3%)
Quintile 5(most)	39,712 (24.2%)
Residential instability	

Quintile 1(least)	17,502 (10.7%)
Quintile 2	25,898 (15.8%)
Quintile 3	29,661 (18.1%)
Quintile 4	34,063 (20.8%)
Quintile 5(most)	56,859 (34.7%)
Ethnic concentration	
Quintile 1(least)	31,974 (19.5%)
Quintile 2	30,886 (18.8%)
Quintile 3	32,080 (19.6%)
Quintile 4	32,647 (19.9%)
Quintile 5(most)	36,396 (22.2%)
Dependency	
Quintile 1(least)	21,609 (13.2%)
Quintile 2	25,374 (15.5%)
Quintile 3	27,847 (17.0%)
Quintile 4	32,319 (19.7%)
Quintile 5(most)	56,834 (34.7%)
Admission type	
Urgent	141,007 (86%)
Elective	22,976 (14%)
Method of entry into the hospital	
Emergency Department (ED)	112,749 (68.8%)
Non-ED[†]	51,234 (31.2%)
Readmission status	
with a history of readmission	8,181(5%)
without a history of readmission	155,802 (95%)
CDCI-admission, mean (SD), median, min-max	1.28 (1.91), 0, 0-18
non-severe multimorbidity status (≤1)	111,465 (68%)
severe multimorbidity status (>1)	52,518 (32%)
ECI-admission, mean (SD), median, min-max	1.11 (1.27), 1, 0-11
non-severe multimorbidity status (≤1)	115,189 (70%)
severe multimorbidity status (>1)	48,794 (30%)
FCI- admission, mean (SD), median, min-max	0.97 (0.92), 1, 0-8
non-severe multimorbidity status (≤1)	122,525 (75%)
severe multimorbidity status (>1)	41,458 (25%)
Acute LOS, mean, SD, median, min-max	12.16 (13.01), 8, 0-347
Outcomes:	
CDCI- discharge, mean (SD), median, min-max	2.14 (2.48), 1, 0-24
non-severe multimorbidity status (≤1)	69,329 (42.3%)
severe multimorbidity status (>1)	62,370 (38%)
ECI- discharge, mean (SD), median, min-max	2.46 (2.40), 2, 0-14
non-severe multimorbidity status (≤1)	56,826 (34.7%)
severe multimorbidity status (>1)	74,873(45.7%)
FCI- discharge, mean (SD), median, min-max	1.87 (1.82), 2, 0-11
non-severe multimorbidity status (≤1)	63,451 (38.7%)
severe multimorbidity status (>1)	68,248 (41.6%)
Death	32,284 (19.7%)

[†] Non-ED refers to the direct admissions from a clinic, doctor's office, or day surgery
* **Note**: CDCI: Charlson Deyo-comorbidity index,
ECI: Elixhauser comorbidity index,
FCI: functional comorbidity index, LOS: length of stay.

4.2. ML-predictability of multimorbidity indices

First, Figure 2 depicts the 10-fold cross-validated AUC of the ML algorithms in predicting the three multimorbidity indices examined in this study. Second, various predictive performance measures on the test set for all utilized algorithms and the three multimorbidity indices are summarized in Table 2. Third, the receiver operating characteristic (ROC) curves for all constructed models are presented in Figure 3 to understand the effects of decision thresholds on the model performance. The results suggest that for all ML algorithms and regardless of the predictive performance criteria, the CDCI can be predicted better than the ECI and the FCI, whereas the FCI is the least predictable index. In addition, XGB and RF are the best performing ML algorithms to predict multimorbidity status.

Fourth, we conducted hypothesis testing to determine whether the differences between the predictive performance of the constructed models are statistically significant (for all multimorbidity indices). The results (available in Appendix E) suggest that all models perform statistically significantly different (at 5% level) for all indices in terms of AUC. However, for the CDCI measure, CART and RF do not perform statistically significantly different in terms of recall.

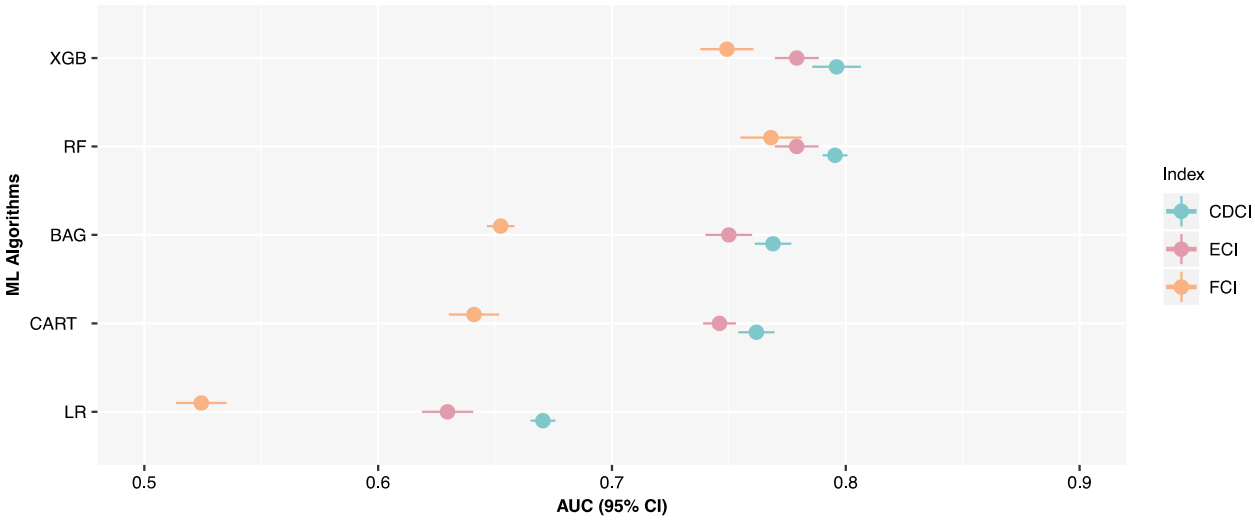


Figure 2. Ten-fold cross-validated AUC of ML algorithms with multimorbidity indices

Note: AUC: area under the ROC curve, BAG: bagging, CDCI: Charlson Deyo-comorbidity index, CART: classification and regression tree, ECI: Elixhauser comorbidity index, FCI: Functional comorbidity index, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting.

Table 2. Out-of-sample predictive performance[†] of ML algorithms with multimorbidity indices

Algorithm	Index	AUC	Accuracy	Recall	Precision	F1
LR	**CDCI**	67.21	64.00	62.60	78.03	69.47
	ECI	63.19	60.00	54.72	38.34	45.00
	FCI	52.19	52.74	52.48	28.16	36.66
BAG	**CDCI**	77.00	67.92	68.80	70.80	69.80
	ECI	75.00	66.50	61.70	57.40	59.40
	FCI	65.23	58.80	58.00	54.96	56.46
CART	**CDCI**	76.70	70.00	69.06	77.69	73.12
	ECI	74.37	68.17	65.64	53.62	59.03
	FCI	64.17	58.80	59.34	48.66	53.47
XGB	**CDCI**	78.78	70.30	**69.27**[‡]	78.43	73.40
	ECI	77.04	68.90	65.90	56.60	60.90
	FCI	64.67	58.50	58.60	49.10	53.47
RF	**CDCI**	**79.55**[‡]	**70.98**[‡]	68.89	**81.37**[‡]	**74.61**[‡]
	ECI	77.70	69.50	49.40	70.51	58.00
	FCI	51.00	50.20	49.00	60.40	54.00

[†] All scores presented in percentage (%).

[‡] The boldfaced numbers indicate the highest value of a predictive performance indicator

Note: AUC: area under the curve, BAG: bagging, CDCI: Charlson Deyo-comorbidity index, CART: Classification and regression tree, ECI: Elixhauser comorbidity index, FCI: Functional comorbidity index, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting

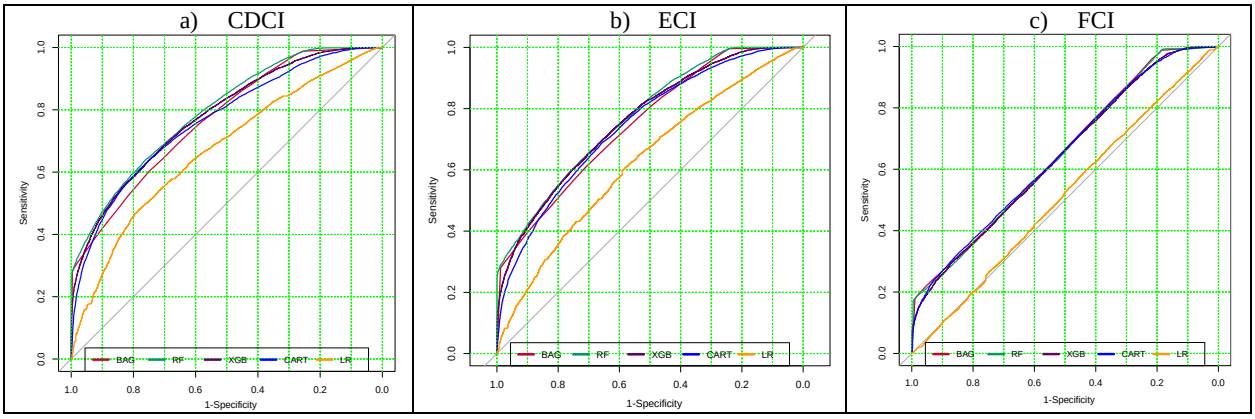


Figure 3. ROC curves for the constructed models for three multimorbidity indices
Note: BAG: bagging, CART: classification and regression tree, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting

Fifth, to assess the overall agreements regarding the multimorbidity status predictions among the three indices (irrespective of their accuracy of the prediction), we calculated Cohen's kappa statistics [61]. Interestingly, the results (Table 3) suggest a *moderate* agreement between CDCI and ECI, a *fair* agreement between ECI and FCI, and a *slight* agreement between CDCI and FCI.

Table 3. Kappa agreements score for multimorbidity indices

	CDCI	ECI	FCI
CDCI	1	0.43 (95% CI: (0.423 – 0.444)[§]	0.13 (95% CI: 0.122 – 0.142)[‡]
ECI		1	0.26 (95% CI: 0.255 – 0.274)[†]
FCI			1

[§] moderate agreement [‡] slight agreement [†] fair agreement
Note: CDCI: Charlson Deyo-comorbidity index, ECI: Elixhauser comorbidity index, FCI: functional comorbidity index.

Sixth, to assess the agreement between the predicted and estimated probability (i.e., risk) of the outcome, we examined the calibration curve for all constructed ML algorithms. Interestingly, the results (available in Appendix F) indicate that XGB provides the highest calibration, while LR was poorly calibrated. These results are consistent with the discrimination power of the algorithms, providing even greater support for the superiority of XGB.

Seventh, we compared the net benefit [59] for all constructed models with two clinical strategies, i.e., (i) "treat all" and (ii) "treat none." The results (available in Appendix G) indicate

that using ML is always preferable to doing nothing (i.e., "treat none"). Quite reasonably, around the starting point (i.e., threshold probability = 0), where there is a low tolerance for risk, "treat all" provides the same benefit as ML. Otherwise, the net benefits of most examined ML algorithms (except LR) clearly dominate that of "treat all," with the XGB and RF exhibiting the highest net benefits. In summary, using ML is beneficial for all threshold probabilities (never dominated by the two strategies), and the benefits are associated with the predictive performance of the ML algorithms.

Eighth, to assess the differences between population groups in terms of predictive performance, we conducted a group analysis based on gender. The results (available in Appendix H) indicate a slight difference in terms of the predictive performance among ML algorithms for males versus females. Accordingly, sensitivity and F1-measure were higher for males compared to females, while the reverse holds for specificity.

Finally, to investigate the impact of the removed missing values (0.3% missingness pertaining to the marginalization variables only) on the performance of the constructed models, we conducted a sensitivity analysis by rerunning all MLs after imputing the missing values with the median of the existing observations. The results indicated that imputing the missing values did not change the predictive performance of the ML algorithms.

4.3. Prognostication of the multimorbidity indices

To assess the prognostic value of the multimorbidity indices in predicting patient-important outcomes, we explored the association between the multimorbidity indices and important adverse events (i.e., mortality and readmission within 30 days post-admission). Figure 4 depicts the results. Accordingly, an increased CDCI is strongly associated with the increase in both adverse events

($RRR^{CDCI}_{mortality}$ = 3.44, 95% CI = 3.21–3.68, p-value <0.001; $RRR^{CDCI}_{readmission}$ = 1.36, 95% CI = 1.31–1.40, p-value <0.001). An increased ECI is also strongly associated with an increase in both events, although with smaller magnitudes than CDCI ($RRR^{ECI}_{mortality}$ = 2.00, 95% CI = 1.87–2.15, p-value <0.001; $RRR^{ECI}_{readmission}$ = 1.24, 95% CI = 1.20–1.29, p-value <0.001). While an increased FCI is strongly associated with increased readmission ($RRR^{FCI}_{readmission}$=1.27, 95% CI = 1.22–1.31, p-value <0.001), it reduces mortality *relative to observing no-event* ($RRR^{FCI}_{mortality}$ = 0.83, 95% CI = 0.77–0.91, p-value <0.001).

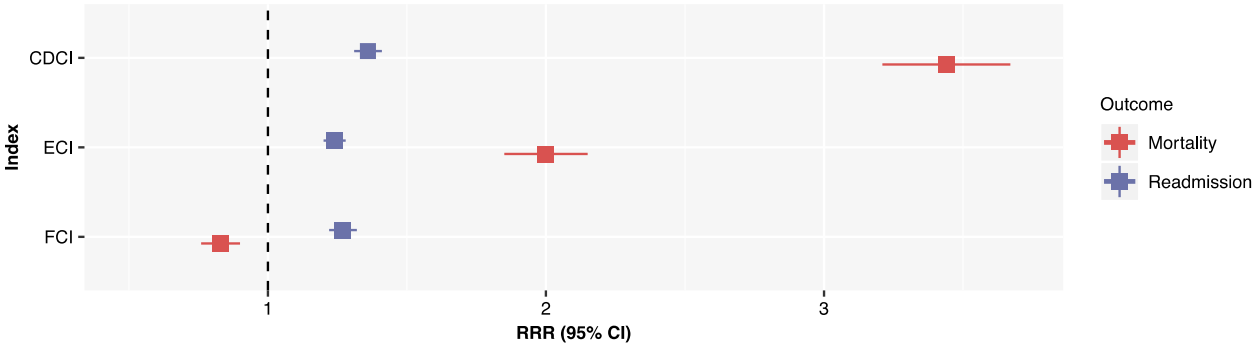


Figure 4. Association of multimorbidity indices with outcomes

Figure 5 summarizes the results of the two analytics in our study. Remarkably, the most predictable multimorbidity index (i.e., CDCI) also offers the highest prognostications regarding adverse events.

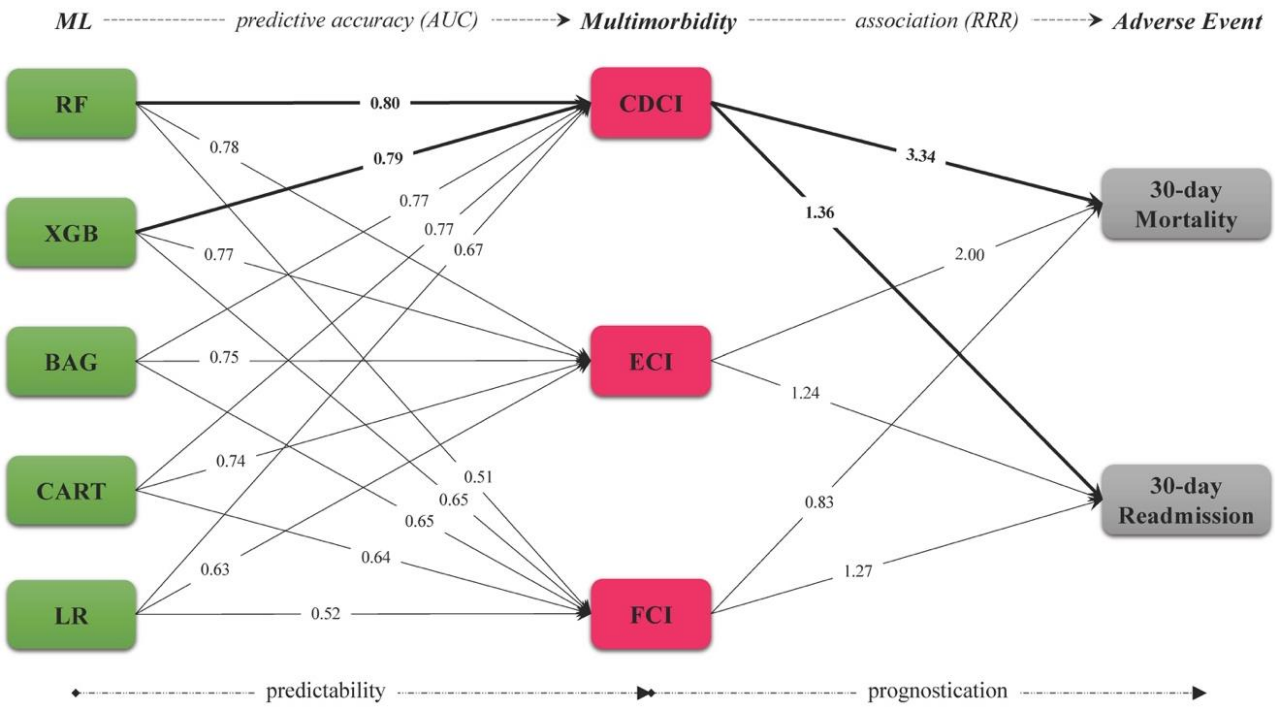


Figure 5. Summary of results regarding the predictability and prognostication of multimorbidity indices

5. Discussion

5.1. Principal findings

Using large longitudinal data on older delayed-discharge patients, this two-stage study evidenced the ML-predictability of patient complexity based on multimorbidity indices and their prognostication significance in predicting patient-important outcomes, such as mortality and readmission. We provided first-hand evidence on this topic through robust comparisons of several predictive accuracy measures using multiple ML algorithms with three common multimorbidity indices. Our results indicate that regardless of the type of the ML algorithm or the measure for predictive performance assessment, the CDCI is the most predictable index, and FCI is the least predictable. Interestingly, these two indices had the least agreement regarding their predictions.

More remarkably, the prognostication analytics in our study revealed that the most predictable index (i.e., CDCI) also has the greatest strength in predicting adverse events.

The differences between multimorbidity indices (in terms of both predictability and prognostication) as well as their degree of prediction agreements may stem from the type of comorbidity items included in the index and the way the items are combined. Although the CDCI consists of fewer comorbidities than the ECI and FCI do, the combination of the comorbidity items in the CDCI is weighted based on their severity to account for the disease burden, which contrasts with the simple summation of the comorbidity items in the ECI and FCI. Several studies have reported the superior prognostic performance of weighted measures of multimorbidity compared with simple measures [41,62–64]. We contribute to this literature by highlighting the importance of accounting for the severity of the diseases for better prognostications with multimorbidity indices, but we are the first to highlight such importance for improving the predictability of multimorbidity indices.

The FCI does not include severe chronic conditions, such as metastatic cancer, dementia, or HIV, which are highly associated with adverse events, particularly mortality [22]. Instead, by including physical functioning-related conditions (e.g., arthritis, hearing and visual impairment, osteoporosis, and degenerative disk disease) — which are better predictors of hospital readmission, particularly among older adults [65,66] — the FCI is essentially a better predictor for hospital readmission than it is for death [23,39]. Consistent with our research, other studies have reported that the CDCI outperforms the ECI in predicting adverse events in the cohort of patients with diabetes [67] and lung cancer [68]. Our results also agree with the literature supporting the superior prognostic significance of the CDCI compared with the FCI [47]. Finally, our findings agree with

previous studies providing evidence that the CDCI and ECI are associated with an increased risk of mortality and readmission [37–44].

5.2. Clinical and policy implications

Because the multimorbidity burden is associated with an increased risk of adverse outcomes, such as readmission and mortality, as well as exponential increases in healthcare costs, healthcare providers should aim to rethink the management of the delayed discharge patients' care needs and implement holistic prevention policies to reduce their risk of adverse outcomes. Recent studies have concluded that patients experience catastrophic health service use and health expenditures when the number of their multimorbid conditions increases [69,70].

Advanced knowledge of multimorbidity status and in-hospital development can be used to support clinical decision-making in the care of older adults awaiting community placement. More specifically, data from prior hospital encounters could be used in conjunction with current medical conditions to determine an overall multimorbidity status of the patient. This, in turn, could be used to flag high-risk patients who may require more intensive monitoring despite readiness for discharge. Given they are ready for discharge, older adults with delayed discharge are often overlooked by hospital staff. Instead, they should be monitored frequently because high rates of multimorbidity and geriatric complexity result in transient health conditions and an increased risk of poor patient outcomes, underscoring the need for monitoring these patients in the hospital. Where possible, consideration should be given to implementing ML algorithms within electronic medical systems to proactively predict multimorbidity status. Our findings are also relevant to policymakers, considering that both mortality and readmission are common metrics used by hospitals to gauge the quality of care provided.

5.3. Strengths and limitations

This study presents robust analytics to examine the predictability and prognostication of three common multimorbidity indices using several ML algorithms and various predictive accuracy measures based on large longitudinal data. However, our study is not without limitations. First, this was a retrospective cohort study with limited control over data collection. Second, we did not include frailty indices, which could complement the examined multimorbidity indices in capturing patient complexity among older adults [13]. Finally, this study focused on the older delayed-discharge patients only; hence the generalizability of our study to other populations is limited.

6. Conclusion

To our knowledge, this is the first study to investigate the ML-predictability of patient complexity thoroughly and concurrently via multimorbidity indices and their prognostic value for predicting patient-important outcomes among older adults or those facing delayed hospital discharge.

Our findings highlight the feasibility and utility of predicting multimorbidity status early in the care pathway of older hospitalized patients. Advance knowledge of patient multimorbidity and complexity can support proactive triage and decision-making about staffing and resource allocation, with the goal of optimizing patient outcomes and discharge planning. Our findings suggest there is value in identifying the multimorbidity status of patients, as it is prognostic of the patient-important outcomes, such as 30-day readmission and mortality. System-level decision- and policymakers can leverage this information to facilitate an upstream and informed discharge process by prioritizing complex patients and providing patient-centered care. More specifically, by gauging staffing and resource allocation on measures like multimorbidity, which are known to

be predictive of health needs and outcomes, resource intensity can be more appropriately matched to meet the distinct needs of the patient and the health system.

Future studies could undertake a revised definition of patient complexity based on both multimorbidity and frailty conditions and investigate their predictability and prognostication. Moreover, further research could extend our analyses to the general population or compare older adults with and without discharge delay. Finally, upcoming studies are required to further investigate the association between multimorbidity status and patient outcomes while controlling for the delayed discharge status.

References

[1] A. Manzano-Santaella, From bed-blocking to delayed discharges: precursors and interpretations of a contested concept, Health Serv. Manage. Res. 23 (2010) 121–127.

[2] D. Challis, J. Hughes, C. Xie, D. Jolley, An examination of factors influencing delayed discharge of older people from hospital, Int. J. Geriatr. Psychiatry. 29 (2014) 160–168.

[3] K.S. McGilton, S. Vellani, J. Babineau, J. Bethell, S.E. Bronskill, E. Burr, M. Keatings, J.E. McElhaney, S. McKay, K. Nichol, Understanding transitional care programmes for older adults who experience delayed discharge: a scoping review protocol, BMJ Open. 9 (2019).

[4] F. Landeiro, K. Roberts, A.M. Gray, J. Leal, Delayed hospital discharges of older patients: a systematic review on prevalence and costs, The Gerontologist. 59 (2019) e86–e97.

[5] A.D. Bai, C. Dai, S. Srivastava, C.A. Smith, S.S. Gill, Risk factors, costs and complications of delayed hospital discharge from internal medicine wards at a Canadian academic medical centre: retrospective cohort study, BMC Health Serv. Res. 19 (2019) 935.

[6] A. Rojas-García, S. Turner, E. Pizzo, E. Hudson, J. Thomas, R. Raine, Impact and experiences of delayed discharge: A mixed-studies systematic review, Health Expect. Int. J. Public Particip. Health Care Health Policy. 21 (2018) 41–56. https://doi.org/10.1111/hex.12619.

[7] F. Landeiro, K. Roberts, A.M. Gray, J. Leal, Delayed Hospital Discharges of Older Patients: A Systematic Review on Prevalence and Costs, The Gerontologist. 59 (2019) e86–e97. https://doi.org/10.1093/geront/gnx028.

[8] R. McCloskey, P. Jarrett, C. Stewart, P. Nicholson, Alternate level of care patients in hospitals: what does dementia have to do with this?, Can. Geriatr. J. 17 (2014) 88.

[9] N. Zychlinski, A. Mandelbaum, P. Momčilović, I. Cohen, Bed blocking in hospitals due to scarce capacity in geriatric institutions—cost minimization via fluid models, Manuf. Serv. Oper. Manag. 22 (2020) 396–411.

[10] H.H. Fenn, A. Hategan, J.A. Bourgeois, Inpatient Geriatric Psychiatry: Optimum Care, Emerging Limitations, and Realistic Goals, Springer, 2019.

[11] A.A. Uijen, E.H. van de Lisdonk, Multimorbidity in primary care: prevalence and trend over the last 20 years, Eur. J. Gen. Pract. 14 Suppl 1 (2008) 28–32. https://doi.org/10.1080/13814780802436093.

[12] A. Marengoni, S. Angleman, R. Melis, F. Mangialasche, A. Karp, A. Garmen, B. Meinow, L. Fratiglioni, Aging with multimorbidity: a systematic review of the literature, Ageing Res. Rev. 10 (2011) 430–439. https://doi.org/10.1016/j.arr.2011.03.003.

[13] D.L. Vetrano, K. Palmer, A. Marengoni, E. Marzetti, F. Lattanzio, R. Roller-Wirnsberger, L. Lopez Samaniego, L. Rodríguez-Mañas, R. Bernabei, G. Onder, Joint Action ADVANTAGE WP4 Group, Frailty and Multimorbidity: A Systematic Review and Meta-analysis, J. Gerontol. A. Biol. Sci. Med. Sci. 74 (2019) 659–666. https://doi.org/10.1093/gerona/gly110.

[14] A. Hassaine, G. Salimi-Khorshidi, D. Canoy, K. Rahimi, Untangling the complexity of multimorbidity with machine learning, Mech. Ageing Dev. 190 (2020) 111325.

[15] B.S. Müller, L. Uhlmann, P. Ihle, C. Stock, F. von Buedingen, M. Beyer, F.M. Gerlach, R. Perera, J.M. Valderas, P. Glasziou, M. van den Akker, C. Muth, Development and internal validation of prognostic models to predict negative health outcomes in older patients with multimorbidity and polypharmacy in general practice, BMJ Open. 10 (2020) e039747. https://doi.org/10.1136/bmjopen-2020-039747.

[16] C. Bähler, C.A. Huber, B. Brüngger, O. Reich, Multimorbidity, health care utilization and costs in an elderly community-dwelling population: a claims data based observational study, BMC Health Serv. Res. 15 (2015) 1–12.

[17] K. Moffat, S.W. Mercer, Challenges of managing people with multimorbidity in today's healthcare systems, BMC Fam. Pract. 16 (2015) 1–3.

[18] C. Liddy, V. Blazkho, K. Mill, Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature, Can. Fam. Physician. 60 (2014) 1123–1133.

[19] M.L. Schiøtz, D. Høst, A. Frølich, Involving patients with multimorbidity in service planning: perspectives on continuity and care coordination, J. Comorbidity. 6 (2016) 95–102.

[20] S.M. McPhail, Multimorbidity in chronic disease: impact on health care resources and costs, Risk Manag. Healthc. Policy. 9 (2016) 143.
[21] R.A. Deyo, D.C. Cherkin, M.A. Ciol, Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases, J. Clin. Epidemiol. 45 (1992) 613–619.
[22] A. Elixhauser, C. Steiner, D.R. Harris, R.M. Coffey, Comorbidity measures for use with administrative data, Med. Care. 36 (1998) 8–27. https://doi.org/10.1097/00005650-199801000-00004.
[23] D.L. Groll, T. To, C. Bombardier, J.G. Wright, The development of a comorbidity index with physical function as the outcome, J. Clin. Epidemiol. 58 (2005) 595–602.
[24] C. Bähler, C.A. Huber, B. Brüngger, O. Reich, Multimorbidity, health care utilization and costs in an elderly community-dwelling population: a claims data based observational study, BMC Health Serv. Res. 15 (2015) 23. https://doi.org/10.1186/s12913-015-0698-2.
[25] L. Han, A. Clegg, T. Doran, L. Fraser, The impact of frailty on healthcare resource use: a longitudinal analysis using the Clinical Practice Research Datalink in England, Age Ageing. 48 (2019) 665–671. https://doi.org/10.1093/ageing/afz088.
[26] F. Mowbray, A.-A. Brousseau, E. Mercier, D. Melady, M. Émond, A.P. Costa, Examining the relationship between triage acuity and frailty to inform the care of older emergency department patients: Findings from a large Canadian multisite cohort study, Can. J. Emerg. Med. 22 (2020) 74–81. https://doi.org/10.1017/cem.2019.432.
[27] P.E. Lønning, Breast cancer prognostication and prediction: are we making progress?, Ann. Oncol. 18 (2007) viii3–viii7.
[28] J.M. Sutherland, R.T. Crump, Alternative level of care: Canada's hospital beds, the evidence and options, Healthc. Policy. 9 (2013) 26.
[29] A.P. Costa, J.W. Poss, T. Peirce, J.P. Hirdes, Acute care inpatients with long-term delayed-discharge: evidence from a Canadian health region, BMC Health Serv. Res. 12 (2012) 172.
[30] R. McCloskey, P. Jarrett, C. Stewart, P. Nicholson, Alternate level of care patients in hospitals: what does dementia have to do with this?, Can. Geriatr. J. 17 (2014) 88.
[31] C. Amy, B. Zagorski, V. Chan, D. Parsons, R. Vander Laan, A. Colantonio, Acute care alternate-level-of-care days due to delayed discharge for traumatic and non-traumatic brain injuries, Healthc. Policy. 7 (2012) 41.
[32] E.A. Spencer, G.A. Ford, M.S. Chan, R. Perera, C. Heneghan, Biomarkers in the prediction of multimorbidity: scoping review, MedRxiv. (2020).
[33] L.T. Mounce, J.L. Campbell, W.E. Henley, M.C.T. Arreal, I. Porter, J.M. Valderas, Predicting incident multimorbidity, Ann. Fam. Med. 16 (2018) 322–329.
[34] K. Wikström, J. Lindström, K. Harald, M. Peltonen, T. Laatikainen, Clinical and lifestyle-related risk factors for incident multimorbidity: 10-year follow-up of Finnish population-based cohorts 1982–2012, Eur. J. Intern. Med. 26 (2015) 211–216.
[35] L. Poole, A. Steptoe, Depressive symptoms predict incident chronic disease burden 10 years later: findings from the English Longitudinal Study of Ageing (ELSA), J. Psychosom. Res. 113 (2018) 30–36.
[36] X. Shang, W. Peng, J. Wu, M. He, L. Zhang, Leading determinants for multimorbidity in middle-aged Australian men and women: A nine-year follow-up cohort study, Prev. Med. 141 (2020) 106260.
[37] S. Marventano, G. Grosso, A. Mistretta, M. Bogusz-Czerniewicz, R. Ferranti, F. Nolfo, G. Giorgianni, S. Rametta, F. Drago, F. Basile, Evaluation of four comorbidity indices and Charlson comorbidity index adjustment for colorectal cancer patients, Int. J. Colorectal Dis. 29 (2014) 1159–1169.
[38] L. Azzalini, M. Chabot-Blanchet, D.A. Southern, A. Nozza, S.B. Wilton, M.M. Graham, G.M. Gravel, J.-P. Bluteau, J.-L. Rouleau, M.-C. Guertin, A disease-specific comorbidity index for predicting mortality in patients admitted to hospital with a cardiac condition, CMAJ. 191 (2019) E299–E307.
[39] A. Kumar, A.M. Karmarkar, J.E. Graham, L. Resnik, A. Tan, A. Deutsch, K.J. Ottenbacher, Comorbidity Indices Versus Function as Potential Predictors of 30-Day Readmission in Older Patients Following Postacute Rehabilitation, J. Gerontol. A. Biol. Sci. Med. Sci. 72 (2017) 223–228. https://doi.org/10.1093/gerona/glw148.
[40] Z.-W. Zhan, Y.-A. Chen, Y.-H. Dong, Comparative Performance of Comorbidity Measures in Predicting Health Outcomes in Patients with Chronic Obstructive Pulmonary Disease, Int. J. Chron. Obstruct. Pulmon. Dis. 15 (2020) 335–344. https://doi.org/10.2147/COPD.S229646.

[41] N. Sharma, R. Schwendimann, O. Endrich, D. Ausserhofer, M. Simon, Comparing Charlson and Elixhauser comorbidity indices with different weightings to predict in-hospital mortality: an analysis of national inpatient data, BMC Health Serv. Res. 21 (2021) 13. https://doi.org/10.1186/s12913-020-05999-5.

[42] M. Cai, E. Liu, R. Zhang, X. Lin, S.E. Rigdon, Z. Qian, R. Belue, J.-J. Chang, Comparing the Performance of Charlson and Elixhauser Comorbidity Indices to Predict In-Hospital Mortality Among a Chinese Population, Clin. Epidemiol. 12 (2020) 307–316. https://doi.org/10.2147/CLEP.S241610.

[43] H.-J. Chang, P.-C. Chen, C.-C. Yang, Y.-C. Su, C.-C. Lee, Comparison of Elixhauser and Charlson Methods for Predicting Oral Cancer Survival, Medicine (Baltimore). 95 (2016) e2861. https://doi.org/10.1097/MD.0000000000002861.

[44] K.S. Ladha, K. Zhao, S.A. Quraishi, T. Kurth, M. Eikermann, H.M.A. Kaafarani, E.N. Klein, R. Seethala, J. Lee, The Deyo-Charlson and Elixhauser-van Walraven Comorbidity Indices as predictors of mortality in critically ill patients, BMJ Open. 5 (2015) e008990. https://doi.org/10.1136/bmjopen-2015-008990.

[45] I. Martín-Lesende, E. Recalde, P. Viviane-Wunderling, T. Pinar, F. Borghesi, T. Aguirre, M. Recio, M.E. Martinez, J. Asua, Mortality in a cohort of complex patients with chronic illnesses and multimorbidity: a descriptive longitudinal study, BMC Palliat. Care. 15 (2016) 1–9.

[46] Matheson FI, van Ingen, 2016 Ontario marginalization index, Tor. St Michael's Hosp. 2018 Jt. Publ. Public Health Ont. (2018).

[47] P. Putrik, S. Ramiro, E. Lie, K. Michaud, M.K. Kvamme, A.P. Keszei, T.K. Kvien, T. Uhlig, A. Boonen, Deriving common comorbidity indices from the MedDRA classification and exploring their performance on key outcomes in patients with rheumatoid arthritis, Rheumatology. 57 (2018) 548–554.

[48] L.B. Goldstein, G.P. Samsa, D.B. Matchar, R.D. Horner, Charlson Index comorbidity adjustment for ischemic stroke outcome studies, Stroke. 35 (2004) 1941–1945.

[49] A. De Hond, W. Raven, L. Schinkelshoek, M. Gaakeer, E. Ter Avest, O. Sir, H. Lameijer, R.A. Hessels, R. Reijnen, E. De Jonge, Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope?, Int. J. Med. Inf. 152 (2021) 104496.

[50] S.-F. Sung, L.-C. Hung, Y.-H. Hu, Developing a stroke alert trigger for clinical decision support at emergency triage using machine learning, Int. J. Med. Inf. 152 (2021) 104505.

[51] H. Jiang, H. Mao, H. Lu, P. Lin, W. Garry, H. Lu, G. Yang, T.H. Rainer, X. Chen, Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease, Int. J. Med. Inf. 145 (2021) 104326.

[52] P.O. Fernandes, D.M. Martins, A. de Souza Bozzi, J.P.A. Martins, A.H. de Moraes, V.G. Maltarollo, Molecular insights on ABL kinase activation using tree-based machine learning models and molecular docking, Mol. Divers. (2021) 1–14.

[53] A. Suresh, R. Udendhran, M. Balamurgan, Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers, Soft Comput. 24 (2020) 7947–7953.

[54] X. Song, X. Liu, F. Liu, C. Wang, Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis, Int. J. Med. Inf. (2021) 104484.

[55] T.E. Cowling, D.A. Cromwell, A. Bellot, L.D. Sharples, J. van der Meulen, Logistic regression and machine learning predicted patient mortality from large sets of diagnosis codes comparably, J. Clin. Epidemiol. 133 (2021) 43–52.

[56] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, Chemom. Intell. Lab. Syst. 174 (2018) 33–44.

[57] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (2009) 427–437.

[58] B. Van Calster, L. Wynants, J.F. Verbeek, J.Y. Verbakel, E. Christodoulou, A.J. Vickers, M.J. Roobol, E.W. Steyerberg, Reporting and interpreting decision curve analysis: a guide for investigators, Eur. Urol. 74 (2018) 796–804.

[59] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med. Decis. Making. 26 (2006) 565–574.

[60] K.F. Kerr, M.D. Brown, K. Zhu, H. Janes, Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use, J. Clin. Oncol. 34 (2016) 2534.

[61] P.F. Watson, A. Petrie, Method agreement analysis: a review of correct methodology, Theriogenology. 73 (2010) 1167–1179.
[62] N.R. Thompson, Y. Fan, J.E. Dalton, L. Jehi, B.P. Rosenbaum, S. Vadera, S.D. Griffith, A new Elixhauser-based comorbidity summary measure to predict in-hospital mortality, Med. Care. 53 (2015) 374–379. https://doi.org/10.1097/MLR.0000000000000326.
[63] A.D. Kabboord, D. Godfrey, A.L. Gordon, J.R.F. Gladman, M. Van Eijk, R. van Balen, W.P. Achterberg, The modified functional comorbidity index performed better than the Charlson index and original functional comorbidity index in predicting functional outcome in geriatric rehabilitation: a prospective observational study, BMC Geriatr. 20 (2020) 114. https://doi.org/10.1186/s12877-020-1498-z.
[64] A.D. Kabboord, M. van Eijk, M. Fiocco, R. van Balen, W.P. Achterberg, Assessment of comorbidity burden and its association with functional rehabilitation outcome after stroke or hip fracture: a systematic review and meta-analysis, J. Am. Med. Dir. Assoc. 17 (2016) 1066-e13.
[65] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, S. Kripalani, Risk prediction models for hospital readmission: a systematic review, JAMA. 306 (2011) 1688–1698. https://doi.org/10.1001/jama.2011.1515.
[66] S.R. Greysen, I. Stijacic Cenzer, A.D. Auerbach, K.E. Covinsky, Functional impairment and hospital readmission in Medicare seniors, JAMA Intern. Med. 175 (2015) 559–565. https://doi.org/10.1001/jamainternmed.2014.7756.
[67] J.K. Gurney, J. Stanley, D. Sarfati, The M3 multimorbidity index outperformed both Charlson and Elixhauser indices when predicting adverse outcomes in people with diabetes, J. Clin. Epidemiol. 99 (2018) 144–152. https://doi.org/10.1016/j.jclinepi.2018.04.002.
[68] C.-C. Yang, Y. Fong, L.-C. Lin, J. Que, W.-C. Ting, C.-L. Chang, H.-M. Wu, C.-H. Ho, J.-J. Wang, C.-I. Huang, The age-adjusted Charlson comorbidity index is a better predictor of survival in operated lung cancer patients than the Charlson and Elixhauser comorbidity indices, Eur. J. Cardiothorac. Surg. 53 (2018) 235–240. https://doi.org/10.1093/ejcts/ezx215.
[69] Y. Zhao, R. Atun, B. Oldenburg, B. McPake, S. Tang, S.W. Mercer, T.E. Cowling, G. Sum, V.M. Qin, J.T. Lee, Physical multimorbidity, health service use, and catastrophic health expenditure by socioeconomic groups in China: an analysis of population-based panel data, Lancet Glob. Health. 8 (2020) e840–e849. https://doi.org/10.1016/S2214-109X(20)30127-3.
[70] A. Gruneir, M. Markle-Reid, K. Fisher, H. Reimer, X. Ma, J. Ploeg, Comorbidity burden and health services use in community-living older adults with diabetes mellitus: a retrospective cohort study, Can. J. Diabetes. 40 (2016) 35–42.

Supplementary Appendix

Appendix A. Data eligibility flow diagram

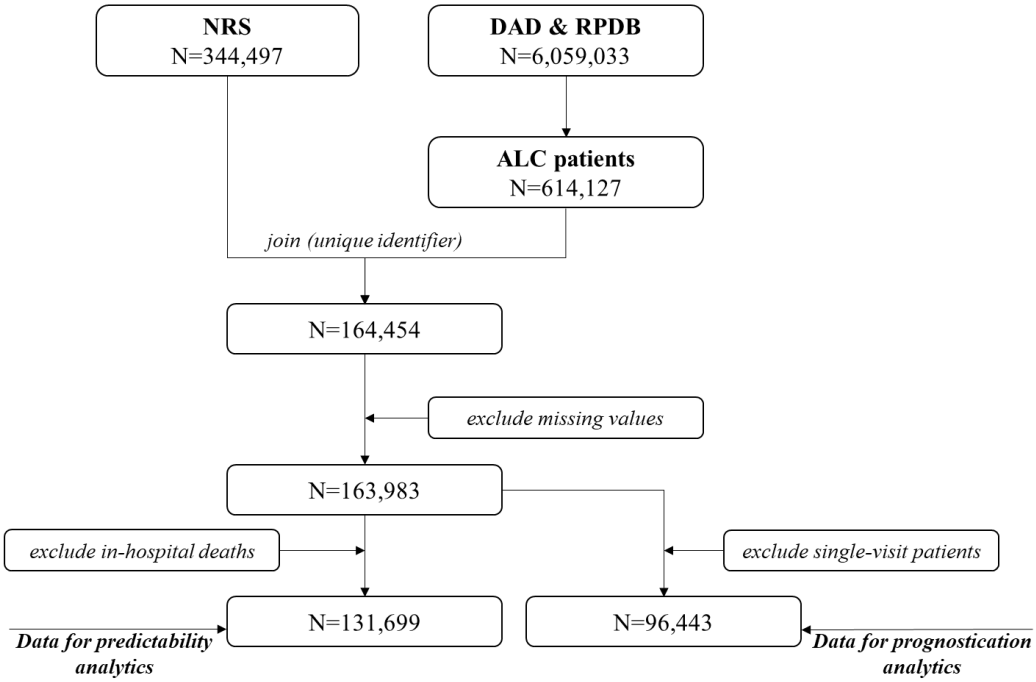


Figure A1. Data eligibility flow diagram

Appendix B. Details of Acute and Chronic Conditions

Table B1. List of acute and chronic conditions included in the multimorbidity indices

Charlson Deyo Comorbidity Index (CDCI)	Elixhauser Comorbidity Index (ECI)	Functional Comorbidity Index (FCI)
1- Myocardial infarction	1. Chronic heart failure	1. Arthritis (rheumatoid and osteoarthritis)
2- Chronic heart failure	2. Cardiac arrhythmia	2. Osteoporosis
3- Peripheral arterial disease	3. Valvular disease	3. Asthma
4- Cerebrovascular disease	4. Pulmonary circulation disorders	4. Chronic obstructive pulmonary disease (COPD), acquired respiratory distress syndrome (ARDS), or emphysema
5- Dementia	5. Peripheral vascular disorders	5. Angina
6- Chronic pulmonary disease	6. Uncomplicated hypertension	6. Congestive heart failure (or heart disease)
7- Connective tissue disease / Rheumatic disease	7. Complicated hypertension	7. Heart attack (myocardial infarct)
8- Peptic ulcer disease	8. Paralysis	8. Neurological disease (such as multiple sclerosis or Parkinson's)
9- Mild liver disease	9. Other neurological disorders	9. Stroke or TIA
10- Uncomplicated diabetes	10. Chronic pulmonary disease	10. Peripheral vascular disease
11- Complicated diabetes	11. Uncomplicated diabetes	11. Diabetes types I and II
12- Paraplegia and hemiplegia	12. Complicated diabetes	12. Upper gastrointestinal disease (ulcer, hernia, reflux).
13- Renal disease	13. Hypothyroidism	13. Depression
14- Cancer	14. Renal failure	14. Anxiety or panic disorders
15- Moderate or severe liver disease	15. Liver disease	15. Visual impairment (such as cataracts, glaucoma, macular degeneration)
16- Metastatic carcinoma	16. Peptic ulcer disease excluding bleeding	16. Hearing Impairment (very hard of hearing, even with hearing aids)
17- AIDS/HIV	17. AIDS/HIV	17. Degenerative disc disease (back disease, spinal stenosis, or severe chronic back pain)
	18. Lymphoma	18. Obesity and/or body mass index>30
	19. Metastatic cancer	
	20. Solid tumor without metastasis	
	21. Rheumatoid arthritis / Collagen vascular disease	
	22. Coagulopathy	
	23. Obesity	
	24. Weight loss	
	25. Fluid and electrolyte disorders	
	26. Blood loss anemia	
	27. Deficiency anemia	
	28. Alcohol abuse	
	29. Drug abuse	
	30. Psychoses	
	31. Depression	

Appendix C. Details of Hyperparameter Tuning

This section provides details of the hyperparameter tuning procedures for the ML algorithms examined in the study.

C.1 Hyperparameter Tuning for The CART Algorithm

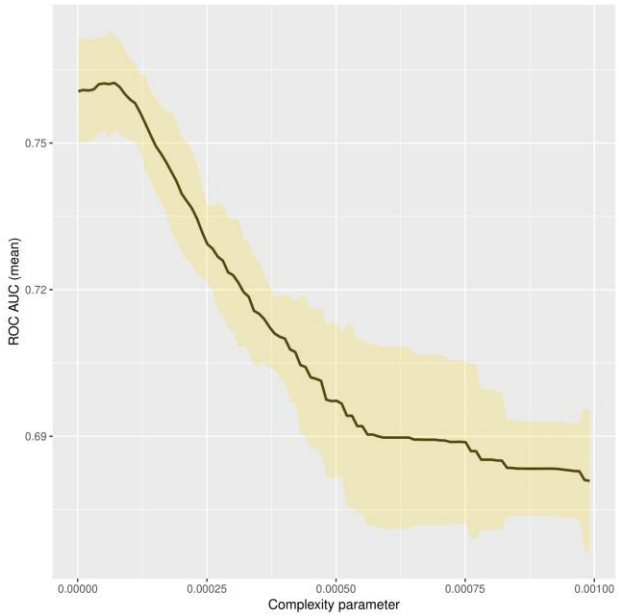


Figure C1. Tuning Complexity Parameter for CART using Grid Search

C.2 Hyperparameter Tuning for The RF Algorithm

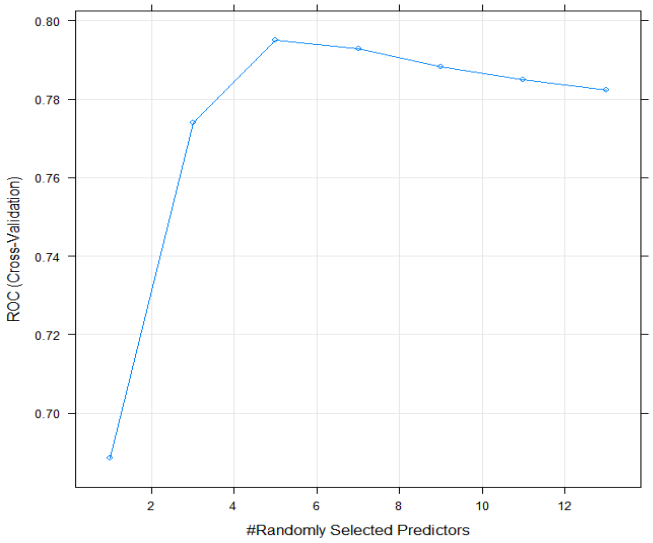


Figure C2. Tuning the number of randomly selected predictors (mtry) for RF using Grid Search

C.3 Hyperparameter Tuning for XGB Algorithm

For tuning the XGB hyperparameters, we applied the grid search hyperparameters tuning technique iteratively to shrink and improve the search range based on the more promising value in the previous iteration. The figures below detail the XGB hyperparameter tuning steps.

C.3.1 Step 1: Tuning Max Tree Depth, Shrinkage/ Learning rate (eta), Iteration

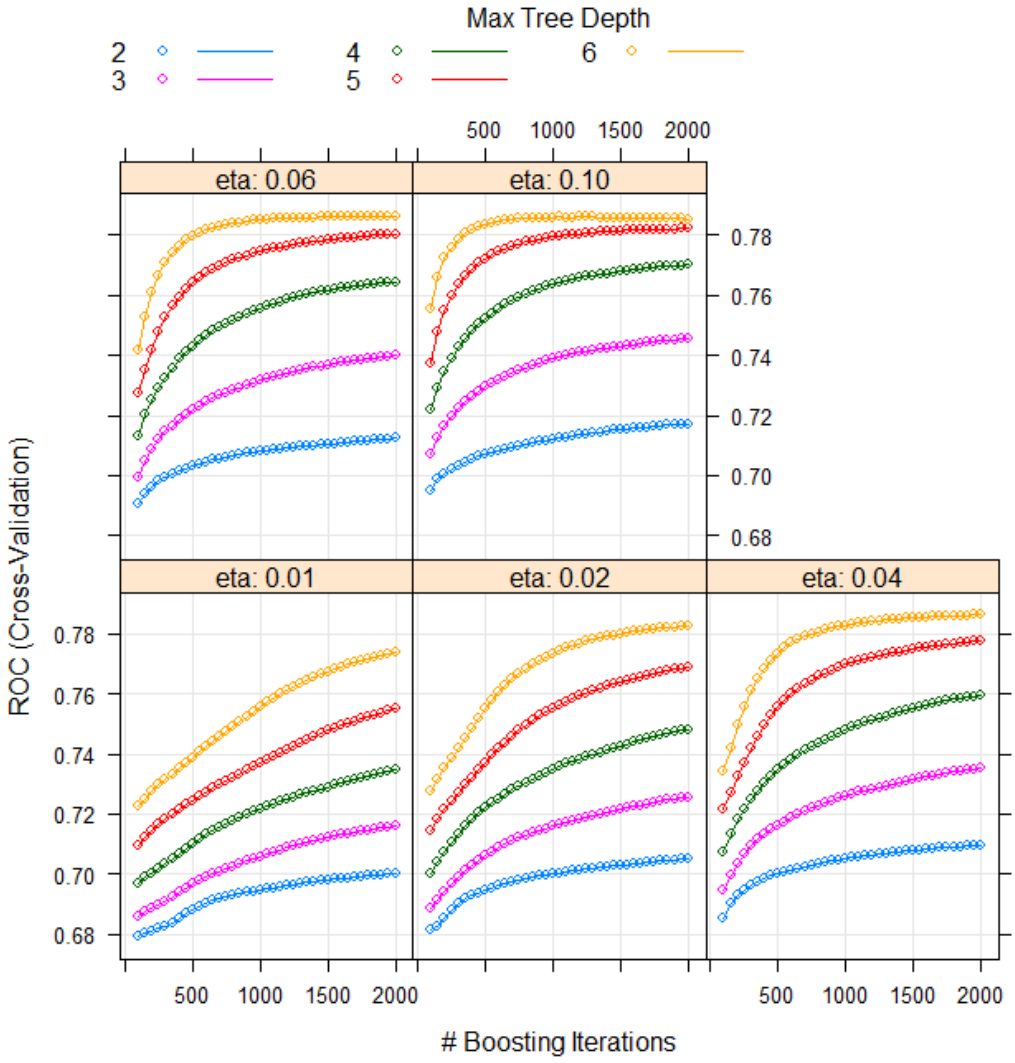


Figure C3. Tuning the max depth tree, shrinkage, and boosting iteration for XGB using Grid Search
**note: Learning rate: {0.01, 0.02, 0.04, 0.06, 0.10}; Max Depth:{2,3,4,5,6}; iteration:100-2000*

C.3.2 Step 2: Tuning Max Tree Depth, Shrinkage/ Learning rate (eta), Iteration, Minimum Sum of Instance Weight

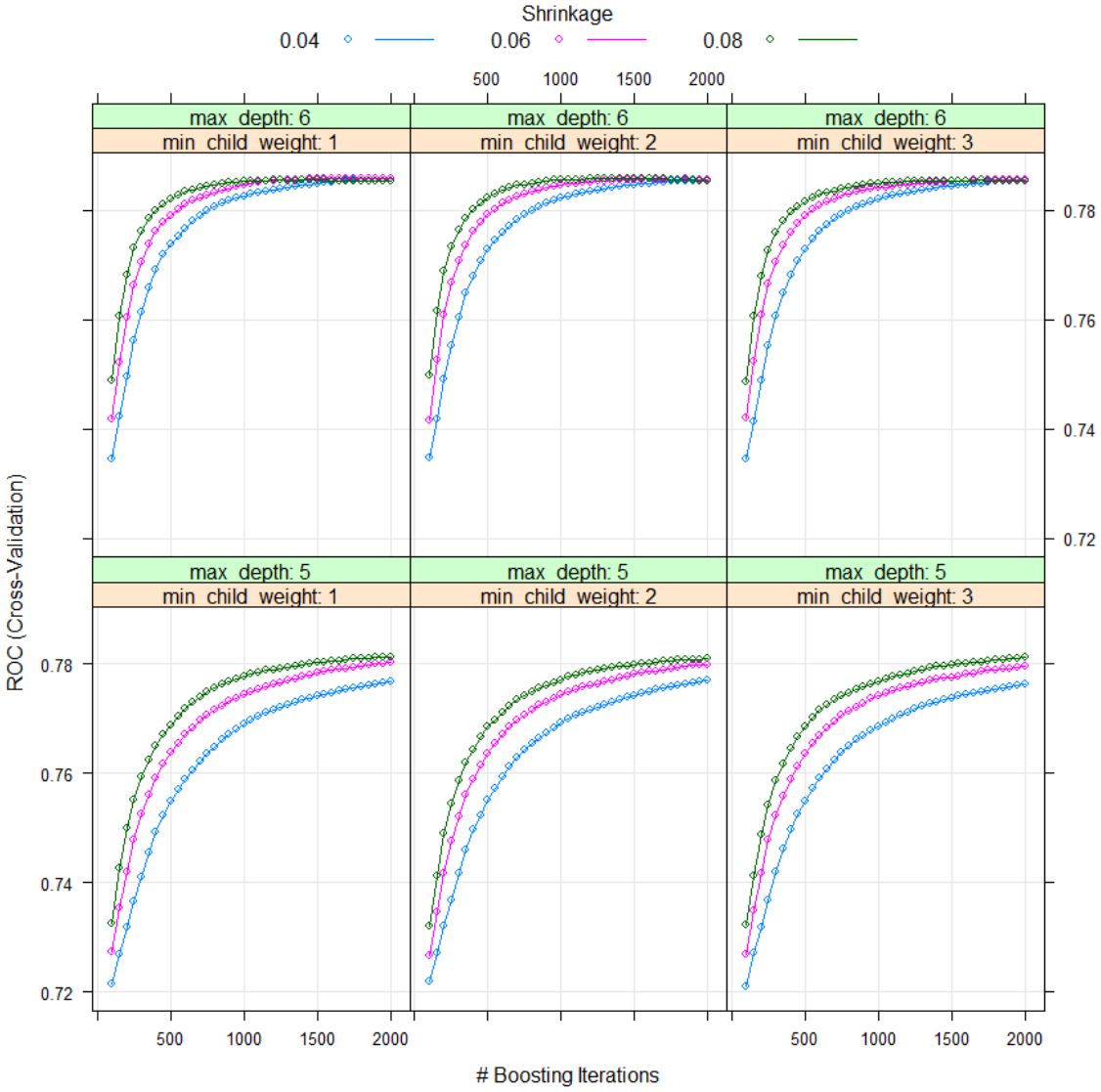


Figure C4. Tuning the max depth tree, shrinkage, boosting iteration, and min sum of instance weight for XGB using Grid Search

**note: Learning rate: {0.04, 0.06, 0.08}; Max Depth: {5,6}; iteration:100-2000; min sum of instance weight: {1,2,3}*

C.3.3 Step 3: Tuning Max Tree Depth, Shrinkage/ Learning rate (eta), Iteration, Minimum Sum of Instance Weight, and Subsample Ratio

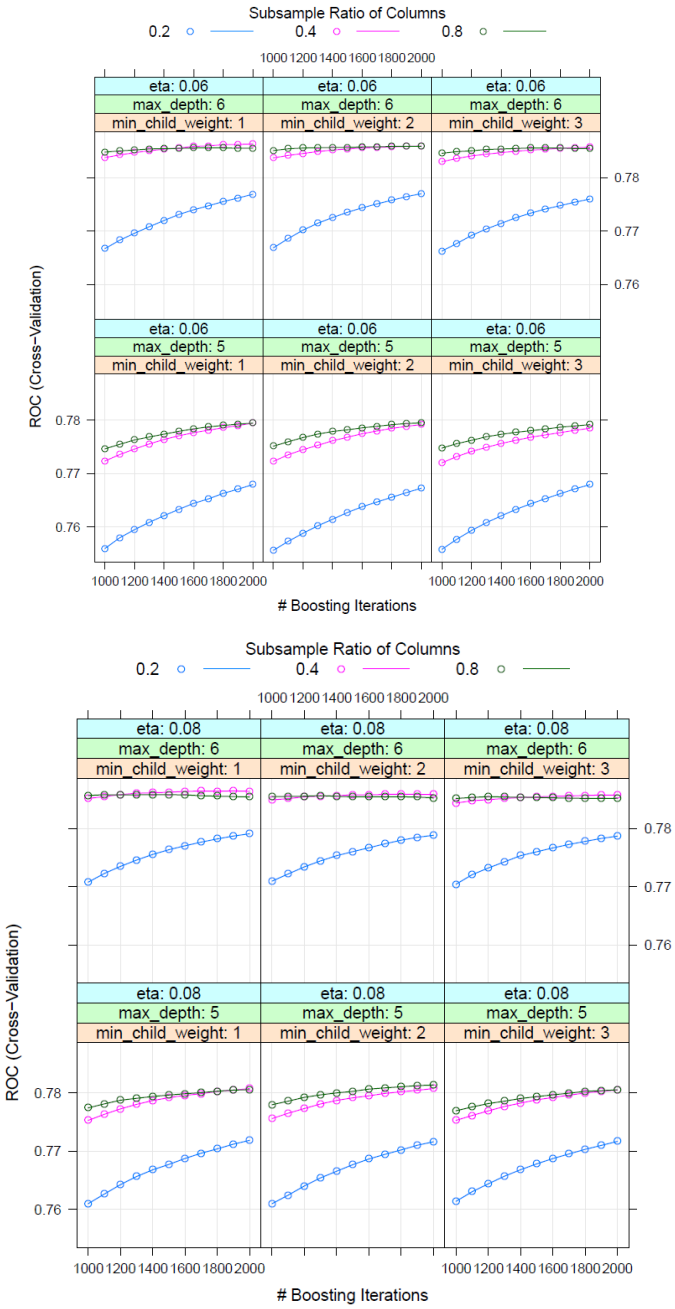


Figure C5. Tuning the max depth tree, shrinkage, boosting iteration, min sum of instance weight, and subsample ratio for XGB using Grid Search

**note: Learning rate: {0.06, 0.08}; Max Depth: {5,6}; iteration:100-2000; min sum of instance weight: {1,2,3}; subsample ratio: {0.2, 0.4, 0.8}*

C.3.4 Step 4: Tuning Max Tree Depth, Shrinkage/ Learning rate (eta), Iteration, Minimum Sum of Instance Weight, Subsample Ratio, Minimum Loss Reduction

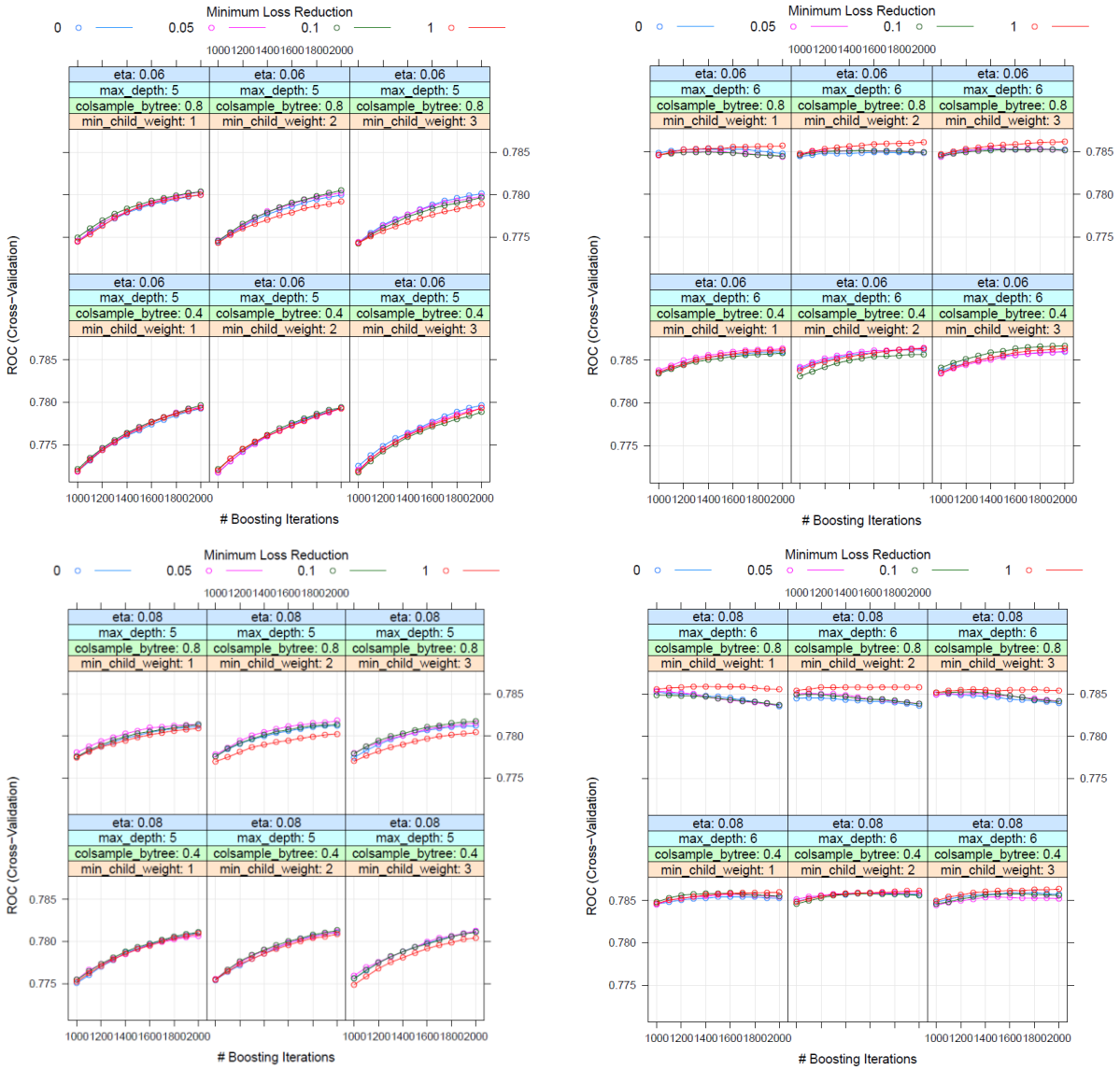


Figure C6. Tuning the max depth tree, shrinkage, boosting iteration, min sum of instance weight, subsample ratio, and min loss reduction for XGB using Grid Search

**note: Learning rate: {0.06, 0.08}; Max Depth:{5,6}; iteration:100-2000; min sum of instance weight: {1,2,3}; subsample ratio: {0.4, 0.8}; min loss reduction: {0,0.05,0.1,1}*

C.3.5 Step 5: Tuning Minimum Loss Reduction, Iteration, Minimum Sum of Instance Weight

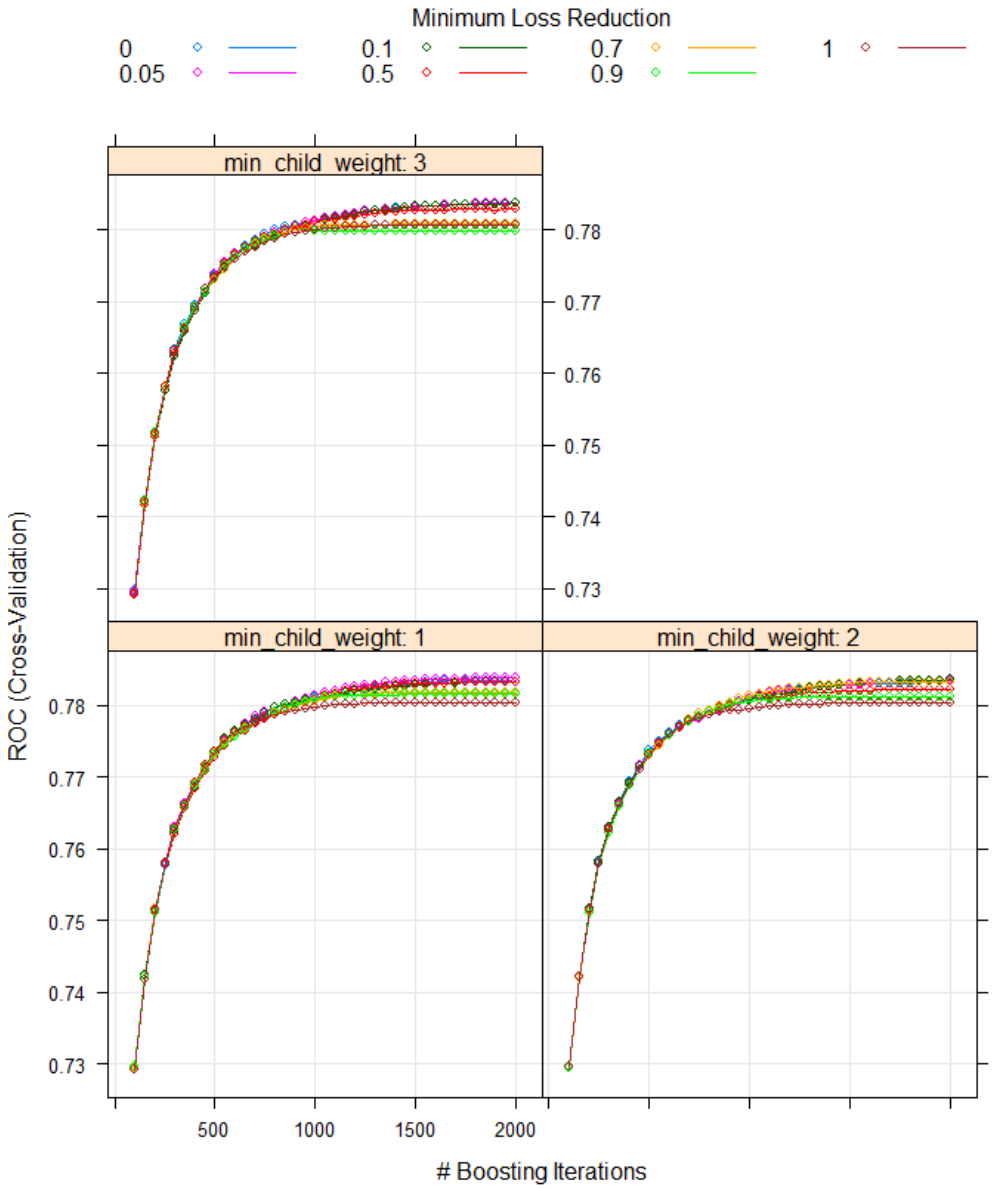


Figure C7. Tuning boosting iteration, min sum of instance weight, and min loss reduction for XGB using Grid Search

**note: iteration:100-2000; min sum of instance weight: {1,2,3}; min loss reduction: {0,0.05,0.1,0.5,0.7,0.9,1}*

C.4 Hyperparameter Tuning Results

Table C1 summarizes hyperparameter tuning results for training the ML algorithms for the predictability analytics.

Table C1. Hyperparameter search space and optimal values for ML algorithms

Algorithm	Hyperparameter Configuration		
	Parameters	**Ranges**	**Optimal Value**
LR.	-		
BAG	-		
CART	cp	1e-6-1e-3 by 0.00001	7.1e-5
RF	mtry	1-13 by 2	5
XGB	NBI	100-2000 by 50	2000
	MTD	2,3,4,5,6	6
	eta	0.01,0.02,0.04,0.06, 0.08, 0.1	0.1
	gamma	0,0.05,0.1,0.5,0.7,0.9,1	1
	SRC	0.4,0.6,0.8,1	0.6
	MSIW	1,2,3	1

Appendix D. Statistical Comparisons of the Multimorbidity Index Means at the Admission and Discharge

Table D1 summarizes the results of hypothesis testing (t-test) regarding the statistical differences between the multimorbidity index means at the admission and discharge.

Table D1. Statistical testing for the difference of index means at the admission and discharge

Index	Mean Adm.	Mean Dis.	diff ‡: 95% CI	p-value
CDCI	1.09	2.14	(-1.06, -1.03)	<0.001
ECI	1.06	2.46	(-1.42, -1.38)	<0.001
FCI	0.97	1.87	(-0.91, -0.88)	<0.001

‡difference = (Index mean at admission – Index mean at discharge)

Note: CDCI: Charlson Deyo-comorbidity index,
ECI: Elixhauser comorbidity index,
FCI: functional comorbidity index
Adm.: admission
Dis.: discharge

Appendix E. Statistical Comparisons of the ML Algorithms' Performances

Tables E1-E3 summarize the results of hypothesis testing regarding the statistical differences between the predictive measures for all ML algorithms and the three multimorbidity indices

examined in the study. The lower diagonal values show the p-values for the statistical test of the difference.

Table E1. Statistical testing to compare the estimated accuracy of ML algorithms (CDCI)

AUC	LR	CART	BAG	RF	XGB
LR					
CART	2.2e-16				
BAG	2.2e-16	2.2e-16			
RF	2.2e-16	2.2e-16	2.2e-16		
XGB	2.2e-16	2.2e-16	1.9e-15	2.2e-16	
Recall	**LR**	**CART**	**BAG**	**RF**	**XGB**
LR					
CART	2.2e-16				
BAG	2.2e-16	2.2e-16			
RF	2.2e-16	0.692	7.7e-10		
XGB	2.2e-16	0.008	2.2e-16	0.0004	

BAG: bagging, CART: classification and regression tree, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting

Table E2. Statistical testing to compare the estimated accuracy of ML algorithms (ECI)

AUC	LR	CART	BAG	RF	XGB
LR					
CART	2.2e-16				
BAG	2.2e-16	2.2e-16			
RF	2.2e-16	2.2e-16	2.2e-16		
XGB	2.2e-16	2.2e-16	1.9e-15	2.2e-16	
Recall	**LR**	**CART**	**BAG**	**RF**	**XGB**
LR					
CART	0.002				
BAG	2.2e-16	1.3e-15			
RF	1.1e-09	7.0e-13	2.2e-16		
XGB	3.2e-07	9.3e-11	2.2e-16	1.2e-05	

BAG: bagging, CART: classification and regression tree, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting

Table E3. Statistical testing to compare the estimated accuracy of ML algorithms (FCI)

AUC	LR	CART	BAG	RF	XGB
LR					
CART	2.2e-16				
BAG	2.2e-16	5.8e-10			
RF	2.2e-16	3.4e-14	0.0003		
XGB	2.2e-16	0.0229	3.58e-07	4.4e-11	
Recall	**LR**	**CART**	**BAG**	**RF**	**XGB**
LR					
CART	2.2e-16				
BAG	2.2e-16	0.0001			
RF	2.2e-16	3.3e-09	3.2e-08		
XGB	2.2e-16	2.2e-16	2.2e-16	0.018	

BAG: bagging, CART: classification and regression tree, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting

Appendix F. Calibration Evaluation for the Constructed Models

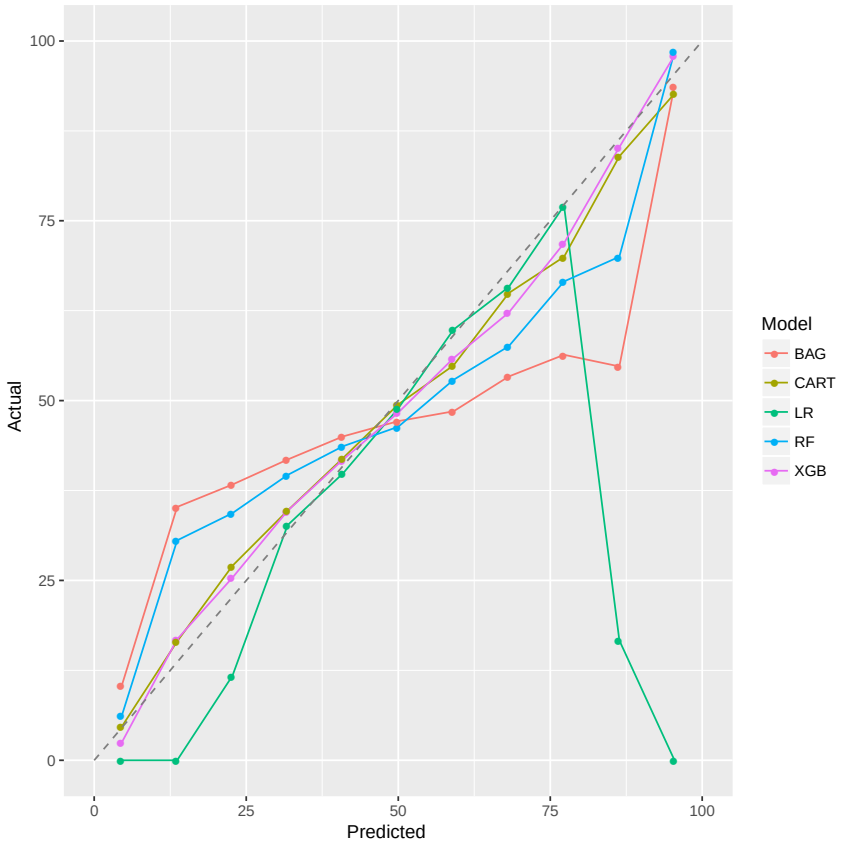


Figure F1. Comparing Calibration of Constructed Models (CDCI)

Appendix G. Decision curve analysis for the Constructed Models

To compare our predictive models in terms of their net benefit at a given threshold probability (P_t), we calculated the standard net benefit for all the examined ML algorithms [1]. The net benefit is equal to the expected benefit to the cases (i.e., true positive rate) minus the expected harm to the controls (i.e., false positive rate multiplied by the threshold probability). For the ease of interpretations, we utilized the standard net benefit, which has a maximum value of 1.0 [2]. In the decision curve analysis, the predictive models are compared to two extreme clinical strategies, assuming that (i) all patients are positive cases, hence choosing "treat all" or (ii) all patients are negative cases, hence "treat none" [1]. In the former case, the true positive rate is equal to the rate of the high-risk event, π, and false positive rate is equal to $1-\pi$. In the latter case, the net benefit is equal to zero [3].

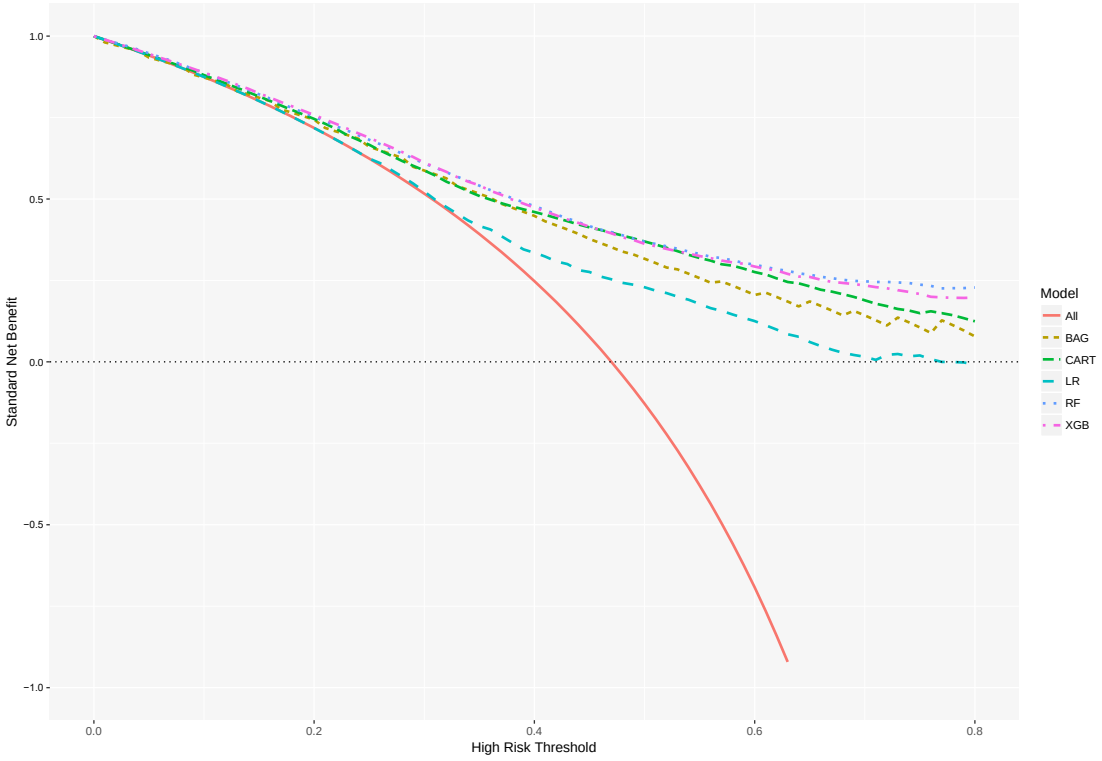


Figure G1. Comparing Standard Net Benefits of the Constructed Models (on CDCI measure)
*Note: horizontal line at zero means net benefit line for treat none strategy, BAG: bagging, CART: classification and regression tree, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting.

Appendix H. Group Analysis: Comparing Model Performance between Genders

Table H1. Comparing the differences in predictive performance for female vs. male XGB (on CDCI measure)

P-Measure	XGB	RF	BAG	CART	LR
MALE					
Accuracy	0.68	0.69	0.67	0.69	0.60
Sensitivity (Recall)	**0.70**	**0.73**	**0.70**	**0.71**	**0.54**
Specificity	**0.66**	**0.64**	**0.64**	**0.67**	0.68
Precision (PPV)	0.71	0.70	0.69	0.71	0.66
F1-measure	**0.70**	**0.71**	**0.69**	**0.71**	**0.59**
FEMALE					
Accuracy	0.71	0.71	0.68	0.71	0.69
Sensitivity (Recall)	**0.51**	**0.52**	**0.58**	0.50	0.71
Specificity	**0.86**	**0.85**	**0.75**	0.87	0.67
Precision (PPV)	0.72	0.72	0.63	0.73	0.71
F1-measure	**0.60**	**0.60**	**0.60**	**0.59**	**0.71**

Note BAG: bagging, CART: classification and regression tree, LR: logistic regression, RF: random forest, XGB: extreme gradient boosting

References

[1] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.

[2] Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. J Clin Oncol 2016;34:2534.

[3] Zhang Z, Rousson V, Lee W-C, Ferdynus C, Chen M, Qian X, et al. Decision curve analysis: a technical note. Ann Transl Med 2018;6.

Chapter 3

Impact of multimorbidity and frailty on adverse outcomes among older delayed discharge patients: implications for healthcare policy

Ghazalbash, S., Zargoush, M., Mowbray, F., Costa, A. (2022). "Impact of multimorbidity and frailty on adverse outcomes among older delayed discharge patients: implications for healthcare policy." **Health Policy.** DOI: https://doi.org/10.1016/j.healthpol.2022.01.004

Preface

The second article in this dissertation, entitled “Impact of multimorbidity and frailty on adverse outcomes among older delayed discharge patients: implications for healthcare policy,” was published in Health Policy Journal (HP) in March 2022. In this research, conceptualization was conducted by S. Ghazalbash and M. Zargoush. Overseen by M. Zargoush, data curation and formal analysis were conducted by S. Ghazalbash, who also drafted the initial version of the manuscript. S. Ghazalbash and M. Zargoush contributed to the investigation and methodology. S. Ghazalbash and M. Zargoush contributed to the interpretation of the results and revised the manuscript for intellectual content. F. Mowbray and A. Costa co-edited the manuscript prior to submission.

Abstract

Objective: To assess the impacts of multiple chronic conditions (MCC) and frailty on 30-day post-discharge readmission and mortality among older patients with delayed discharge.

Data Source/Extraction: We used a retrospective cohort of older patients in the Discharge Abstract Database (DAD) between 2004 and 2017 in Ontario, Canada. We extracted data on patients aged ≥ 65 who experienced delayed discharge during hospitalization (N=353,106).

Study Design: We measured MCC and frailty using the Elixhauser Comorbidity Index (ECI) and the Hospital Frailty Risk Score (HFRS), respectively. We used multinomial logistic regression to model the main and interactive effects of MCC and frailty on the adverse outcomes.

Principal Findings: After adjusting for sex, discharge destination, urban/rural residency, wait time for alternative care, and socioeconomic status, the coexistence of MCC and high frailty increased the relative risk of 30-day mortality and readmission when compared to the references group, i.e., non-MCC patients with low-to-moderate frailty.

Conclusions: Multimorbidity and frailty each provide unique information about adverse outcomes among older patients with delayed discharge but are most informative when examined in unison.

Implications for health policy: To minimize the risk of adverse outcomes among older delayed discharge patients, discharge planning must be tailored to their concurrent multimorbidity and frailty status.

Keywords: multimorbidity; frailty; discharge policy; delayed discharge; hospital readmission; mortality; geriatrics.

1. Introduction

Multimorbidity, also known as multiple chronic conditions (MCC), is present in most older patients and is defined as the coexistence of two or more chronic diseases [1–4]. Frailty, which is also most prevalent in older adults, is a multidimensional syndrome characterized by a heightened vulnerability to adverse health and reduced ability to recover from stressors due to a diminished physiologic reserve [5,6]. These health indicators are associated with patient vulnerability and adverse outcomes [7–9] and provide succinct information regarding overall patient health and complexity. In addition, older adults with MCC are more likely to be frail. However, these health measures can diverge significantly [1], suggesting that they can offer unique prognostic value for determining patient outcomes, clinical decision-making, and disposition planning in hospitalized older patients.

Several studies have focused on the definition of "patient complexity," concluding that multimorbidity is the most commonly used component of patient complexity [10]. However, in the context of the aging population, some studies reported that multimorbidity could not fully capture the complexity of elderly patients [11]. They defined geriatric syndromes, such as frailty, as another dimension of patient complexity [12,13]. Medically complex older adults, i.e., frail patients with multiple chronic conditions, are at greater risk for functional decline, nosocomial injury, medical complications, and death [14–17]. Moreover, complex medical and psychosocial histories complicate discharge planning, resulting in increased adverse outcomes in older adults [18]. Mortality and hospital readmission are commonly examined as hospital quality metrics and patient-important outcomes for older adults [19–21].

The limited availability of community-based services impedes healthcare providers from safely discharging older adults with complex health states, leading to delayed discharge and inefficient use of acute care beds [22]. Delayed discharge is a prevalent global issue in many countries, such as Canada, the United Kingdom, Germany, Norway, Sweden, Scotland, Denmark, France, New Zealand, Australia, and South Korea [23–27]. One-third of hospitalized older adults experience a delayed discharge during hospitalization [28]. In Canada, these patients are designated as Alternate Level of Care (ALC) patients. A delayed discharge patient occupies an acute or post-acute bed despite no longer requiring that intensity of resources and services, leading to the so-called "delayed transfers of care" [29]. Most delayed discharge patients are waiting to be placed in an alternative level of care, such as long-term care (LTC), home care, rehabilitation, or post-acute care, which have limited capacity. Delayed discharge leads to increased healthcare costs, a decline in patient overall health and functional mobility, and an increased rate of adverse medical events, such as infections, falls, and delirium [30–32]. It also has a domino effect on the healthcare system by negatively impacting patient flow across the entire system, which causes emergency department (ED) overcrowding and outpatient elective surgery cancellations [30,31].

These challenges have motivated healthcare managers, policymakers, and discharge planners to investigate evidence-informed strategies to improve discharge policies among older delayed discharge patients [26,33,34]. One barrier to the optimal discharge of delayed discharge patients is an imperfect understanding of their distinct physical and psychosocial needs associated with their complexity [34]. The co-occurrence, severity, and consequences of frailty and multimorbidity are likely greater among delayed discharge patients, highlighting a cohort of mostly complex older

persons following an acute illness [26,35,36]. Advanced knowledge of the effects of multimorbidity and frailty on the adverse outcomes among these patients could help stratify patients at risk for adverse health events and support proactive discharge planning and other interventions both in the hospital and following discharge.

A compelling body of literature has investigated the effects of multimorbidity and frailty on readmission and mortality in the older population. Table S1 in the Appendix presents a detailed account of this literature. However, only a few studies have investigated coexisting multimorbidity and frailty effects on adverse outcomes among older adults [37–39]. The common measures of multimorbidity among these studies included selected chronic condition items [38] or patterns of multimorbidity [37,39]. Frailty was often measured using indices such as Fried's phenotypic criteria [37,39] and the Canadian Study of Health and Aging Clinical Frailty Scale (CFS) [38]. Moreover, these works have focused on general patients [39], community-dwelling older adults [37], or patients with COVID-19 [38]. To our knowledge, no study has investigated the association of multimorbidity and frailty with adverse health outcomes among older adults with delayed discharge.

Through an extensive analysis of large data from a Canadian population, we aimed to examine the coexisting effects of multimorbidity and frailty measured through the Elixhauser Comorbidity Index (ECI) [40,41] and the Hospital Frailty Risk Score (HFRS) [42], respectively, on a series of patient outcomes. The outcomes of interest included 30-day mortality and two means of hospital readmission (via ED or direct) within 30 days post-discharge. The utility of ECI has been validated in the studies that use administrative/clinical data to predict mortality and readmission,

outperforming some other indices, such as the Charlson comorbidity index [43]. Moreover, moving toward using electronic health records (EHR) data to measure frailty indices, the HFRS has shown moderate agreements with other frailty risk scores, such as FI, and correlated well with the 30-day mortality and readmission [44]. Also, this score exhibited a fair agreement with the Fried and Rockwood frailty scales and moderate agreement with the Rockwood Frailty Index [42]. To assess the robustness of our results, we conducted several sensitivity analyses with respect to the changes in 1) the length of outcome window, 2) discharge location, 3) readmission configuration, 4) discharge wait time, and 5) cut-off point for the frailty index dichotomization.

2. Methods

2.1. Study cohort and data sources

We conducted a retrospective cohort study using data extracted from the Discharge Abstract Database (DAD). This database is provided by the Canadian Institute for Health Information (CIHI) housed at the Institute for Clinical Evaluative Sciences (IC/ES), which has been used in several studies [45–48]. All hospitalized patients aged 65 years and older who experienced a delayed discharge between 2004 and 2017 in Ontario, Canada's largest province, were eligible for study inclusion. The delayed discharge designation is made by an appropriate care team that usually comprises physicians, LTC assessors, patient care managers, and discharge planners. This variable is recorded in the DAD by health information management professionals [49]. We used the "ALC" and "delayed discharge" interchangeably throughout this manuscript. This study was approved by the Hamilton Integrated Research Ethics Board (HiREB).

2.2. Population exclusions

For this study, we excluded patients who died during the index hospital visits, were transferred to another service during their hospital stay, left the hospital against medical advice, or had planned hospital visits. When patients had repeated readmissions during the 30-day window, we considered only the first readmission [50,51]. These criteria resulted in a total of 353,106 observations (Fig. S1 in the Appendix).

2.3. Primary independent variables

The primary independent variables were MCC and frailty status. MCC status was assessed through the ECI, a sum of 31 chronic conditions, hence ranging from 0 to 31 [40,41]. ECI is associated with all-cause mortality, length of hospital stay, and healthcare-related costs [52]. For each observation, the disease diagnosis was assessed using the ICD-10 codes (International Statistical Classification of Diseases and Related Health Problems, Tenth Revision) and converted to ECI as instructed by Azzalini and colleagues [41]. We categorized ECI into ECI≥2 and ECI<2, respectively, to designate MCC patients and non-MCC patients (i.e., those with zero or one chronic condition). The dichotomization cut-off of 2 was determined to be the best threshold for discriminating between patients with and without the outcomes. We used the Akaike Information Criterion (AIC) as the goodness of fit measure.

Frailty was measured using the HFRS, which was proposed in a recent Lancet study by Gilbert and colleagues [42]. This measure was developed based on 109 ICD-10 diagnostic codes that were determined to be associated with frailty. Prior work has demonstrated that HFRS is associated with 30-day mortality and readmission [42,48]. For each observation, we calculated the HFRS using the designated ICD-10 diagnostic codes, with each patient allotted a maximum of 25 diagnostic

variables. The R code for calculating the HFRS is available in Appendix E. It is important to note that we measured the HFRS and ECI for all inpatient episodes, rather than considering one randomly-selected admission per patient undertaken by other studies [48]. Our approach requires updating the index calculations after each visit to include all patient episodes. This approach better captures the relationships between longitudinal factors and the outcome and reduces the time-dependent bias explained by van Walraven et al. [53]. In doing this, we followed the procedure recommended by Senot [54]. The details of the ICD10 codes of the items included in the ECI and HFRS indices are available in Appendix B (Table S3). The descriptive analysis indicated a low correlation between these indices (r = 0.17). An HFRS point of 14 was determined to be the best dichotomization cut-off for discriminating between patients with and without the outcomes of interest. The details of our approach for finding the HFRS cut-off point are available in Appendix C. The original study by Gilbert and colleagues categorized the HFRS as low risk (<5), moderate risk (5–15), and high risk (>15) [42]. With the cut-off point of 14, therefore, the frailty measure in our study can be interpreted as *low-to-moderate* (HFRS≤14) versus *high* (HFRS>14).

2.4. Covariates

We adjusted for several covariates, including i) sex, ii) socioeconomic status (marginalization and rural/urban residency), and iii) administrative variables (ALC wait time and discharge destination). Marginalization was measured using the Ontario Marginalization Index (ON-Marg) [55], which consists of material deprivation (a compound measure of income, education, single-parent families, and housing quality), residential instability (a compound measure of dwelling/family characteristics, neighborhood quality, and cohesiveness), ethnic concentration (indicating recent immigration and visible minorities), and dependency (capturing those who are unemployed, unable

to work, and in unpaid professions). These items are measured on a quantile scale from Q1 to Q5, where Q5 represents the most severe level of marginalization. Because the items of marginalization were highly associated with each other, we only used one item at the same time to avoid multicollinearity. In the interest of space and given the importance of racial inequality from health policy standpoints [56,57], we only present the results obtained with the ethnic concentration item. The urban vs. rural residency is defined based on the population size/density and the distance (or travel burden) to an urban center or essential services [58]. ALC wait time (denoted by ALCW) captures the number of days a patient waits, after medical and discharge clearance, to be assigned an alternative care. Finally, the discharge destination was dichotomized based on whether the patient was discharged home with self-care or to an alternative care (e.g., home with support, community care, and LTC). The last two covariates are specific to the delayed discharge patients. Because age was highly correlated with the other covariates, it was not used in the models. Prior studies have determined that frailty, and not age, is predictive of hospital admission and readmission in older adults [8,59].

2.5. Outcomes

The primary outcomes were 30-day post-discharge readmission and mortality. Hospital readmission was defined as any unplanned readmission and was split into readmission via ED and direct readmission to the hospital (e.g., from a clinic, doctor's office, or day surgery). In our data, around 13% of the patients were admitted to the hospital directly. The CIHI has highlighted the importance of considering these two means of readmission for older adults, particularly those with delayed discharge [60]. Past works have also discussed the potential impacts associated with the

direct readmission compared to the readmission via ED (e.g., reducing ED overcrowding and increasing the delays in initial evaluation) [61]. Accordingly, these two modes of admission could lead to different adverse outcomes [62], supporting the idea of separating these two means of readmissions. We assessed the sensitivity of our results to various configurations of the readmission variable. In summary, we defined the 30-day post-discharge outcome variable with four possible levels: mortality (coded as outcome = 3), readmission via ED (outcome = 2), direct readmission (outcome = 1), and neither (outcome = 0). We estimated the relative risk ratio (RRR) with respect to "outcome = 0" (i.e., "no evet") as the reference outcome to facilitate direct comparisons across the outcomes. We also conducted a sensitivity analysis to determine whether the results changed across different outcome window lengths, including 7, 90, 180, and 360 days.

2.6. Statistical analysis

Descriptive statistics were reported using measures of frequency and central tendency. Multinomial logistic regression was performed with interaction analysis to assess the association of frailty and multimorbidity with the outcome variable while controlling for the covariates. Based on the combinations of the ECI and HFRS status, we categorized the MCC-Frailty risk into four levels, which allowed us to investigate the effects of multimorbidity and frailty on the outcomes of interest. These four MCC-Frailty risk categories were:

- **Group 1 (reference; 62%):** having a low-to-moderate level of frailty without MCC (ECI<2, HFRS≤14)
- **Group 2 (3.8%):** having a high level of frailty without MCC (ECI<2, HFRS>14)
- **Group 3 (29.8%):** having a low-to-moderate level of frailty with MCC (ECI≥2, HFRS≤14)
- **Group 4 (4.3%):** having a high level of frailty with MCC (ECI≥2, HFRS>14).

For model simplicity and ease of interpretation, we considered the linear form of interactions. However, we also examined curvilinear interactions between the continuous versions of frailty and multimorbidity indices after adjusting for the potential confounders. We compared the resulting models in terms of AIC in Appendix B (Table S5).

All tests were two-sided with the statistical significance level of α=0.05, and all estimates were presented with their 95% confidence intervals (95% CI). The absence of multicollinearity was verified through the Variance Inflation Factor (VIF). The statistical analyses were performed using R version 3.3.5, packages "*nnet*" and "*stargazer*."

3. Results

The numerical details of all graphs are available in Appendix D.

3.1. Descriptive results regarding patient characteristics

Table S2 in the Appendix displays the patient and hospitalization characteristics in terms of MCC-Frailty risk categories. A total of 353,106 patients aged between 65 and 102 years were included in the study. The mean age of the sample was 78 years, with a standard deviation (SD) of 7.9 years. Most patients were female (59%), hospitalized in an urban setting (88.6%), and discharged to a destination other than home (91%). HFRS scores ranged from 0 to 41, with approximately one-fifth of the patients having a score of zero. ECI scores ranged from 0 to 12, with approximately 37% of the patients having an ECI score of zero. Overall, 16,367 (4.6%) of the patients were readmitted directly, 47,971(13.6%) via ED, and 9,182 (2.6%) died during hospitalization within 30 days after discharge. The mean ALCW was 15.8 days, with an SD of 31.7 days. The missing

values were minimal (with 1.2% missingness pertaining to the marginalization variable only) and were imputed with the median of the existing observations.

3.2. Association of frailty and comorbidity with the outcomes of interest

Fig. 1 displays the adjusted associations (i.e., RRR) between the MCC-Frailty risk categories and the three outcomes compared to the group with the lowest risk (group 1) as the reference group.

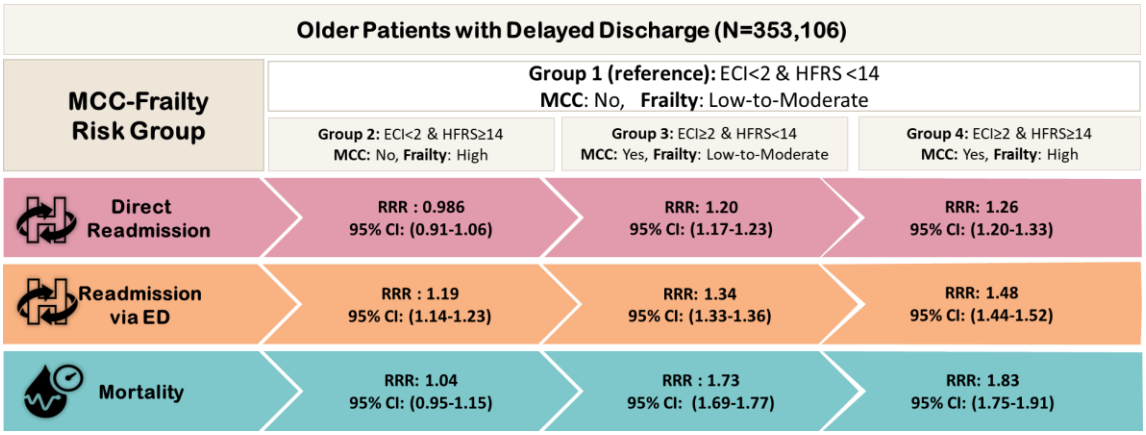


Fig. 1. Adjusted RRR for the MCC-Frailty risk categories
Note: The three outcome events are presented with the same colors across the paper

These results suggest that simultaneously having MCC and a high level of frailty (group 4) increased the relative risk of all three outcomes and had the largest impact on mortality. Similarly, in group 3, mortality had the highest risk, followed by readmission via ED and direct readmission. Moreover, this group's outcomes had a smaller effect size compared to group 4. Having a high level of frailty without MCC (group 2) increased the relative risks of admission via ED, but the associations were not significant for the direct readmission and mortality. As shown in the sensitivity analysis section, this group's risk of death became significant for the windows of length ≥90 days. The adjusted RRR's for all outcomes were similar to the unadjusted results. Our results also indicate that the interaction between multimorbidity and frailty is antagonistic (i.e., the joint

effect is less than the sum of individual effects) for the risks of direct readmission, ED readmission, and mortality. Therefore, failing to account for the interaction between multimorbidity and frailty leads to overestimating the risks of all outcomes.

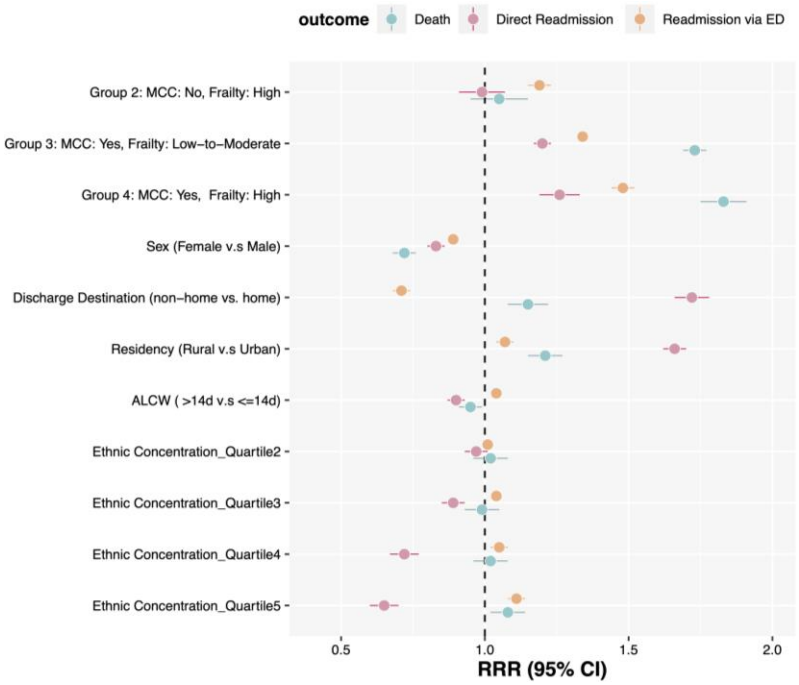


Fig. 2. Detailed results of the adjusted model

The relative risks of all events were lower in females than in males, with the largest difference seen for mortality (RRR=0.72, 95% CI: 0.69 – 0.76). Those who were discharged somewhere other than home had a lower relative risk of readmission via ED (RRR=0.71; 95% CI: 0.68 – 0.74) but a higher relative risk of death (RRR=1.15; 95% CI: 1.09 – 1.22) and direct readmission (RRR=1.72; 95% CI: 1.66 – 1.78) compared to those who were discharged home. Rural residency increased the relative risk of all three events, with the greatest risk pertaining to the direct readmission (RRR=1.66; 95% CI: 1.62 – 1.70) compared to the urban residency. Longer ALC wait time was associated with a lower risk of direct readmission and death but a higher risk of readmission via ED (RRR=1.05; 95% CI: 1.03 – 1.07), although with a small effect size.

Increased marginalization was consistently associated with the decreased risk of direct readmission and the increased risk of readmission via ED and death compared to the group with the lowest degree of marginalization (i.e., Q1). The most marginalized group (Q5: patients from areas with the highest proportion of recent immigrants or visible minorities) had the lowest relative risk of direct readmission (RRR=0.65; 95% CI: 0.61 – 0.70) but the highest relative risk of readmission via ED (RRR=1.11; 95% CI: 1.09 – 1.14) and death (RRR=1.09; 95% CI: 1.02 – 1.14). The results remained relatively unchanged when the ethnic concentration item of the marginalization index was replaced with the other items (i.e., material deprivation, housing instability, and dependency). In the next step, we explored the difference in the outcomes when stratifying the analysis by sex while adjusting for the remaining covariates (Fig. 3). We found apparent differences between males and females, particularly in the risk of death among groups 3 and 4, suggesting clear evidence of an interaction between sex and the MCC-Frailty risk. In other words, the impacts of multimorbidity and frailty on adverse outcomes depend on sex. We did not find any substantial evidence of an interaction between sex and other covariates.

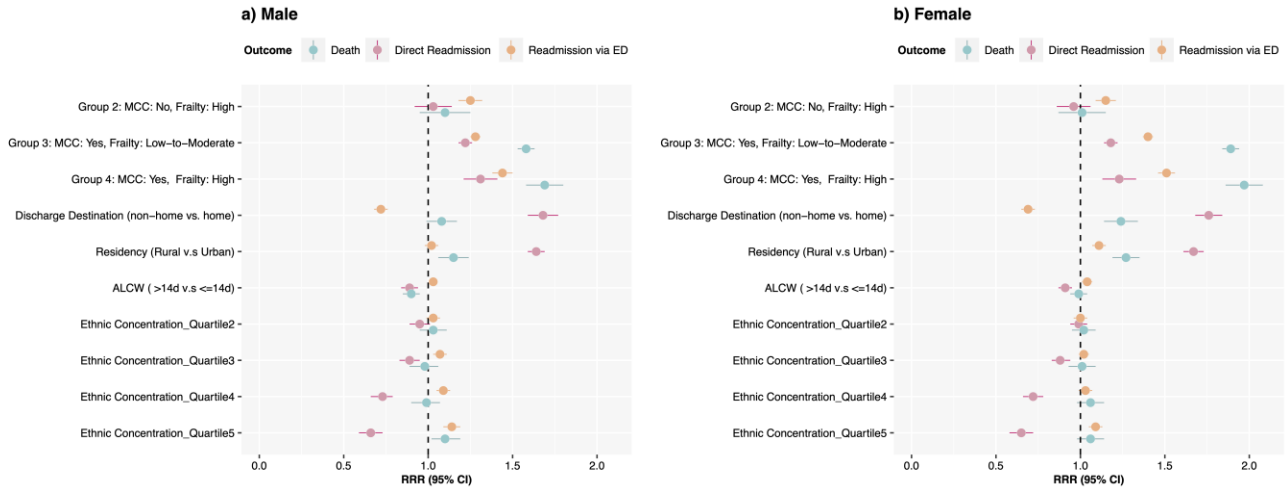


Fig. 3. Results of adjusted model stratified by sex

3.3. Sensitivity analysis

To evaluate the robustness of our findings, we conducted several sensitivity analyses, as follows. First, we altered the primary window length (30 days) to 7, 90, 180, and 360 days to explore if the results change across different window lengths (Fig. 4). In group 2 (Fig. 4-a), the results were more sensitive to the window length before 90 days and became more stable afterward. The risk of death in this group was not statistically significant at 7 and 30 days but became significant for the windows of length ≥90 days. Similarly, in group 3 (Fig. 4-b), the risk of outcomes changed mainly within 90 days and became more stable afterward. In group 4 (Fig. 4-c), all three outcomes demonstrated a strong positive association with the time window after 90 days post-discharge. This observation suggests the time-sensitivity of any interventions (such as early discharge planning), particularly for this group of patients.

Second, to assess the sensitivity of the association between discharge destination and the outcomes, we expanded the alternative care category by separating home care (i.e., "home with support") from other alternative care destinations (i.e., community or LTC). Fig. S2 in Appendix A illustrates the results. One can observe the largest changes in the readmission risks. Unlike the original results (when all types of alternative care were combined, i.e., Fig S2-a), "home with support" decreases the relative risk of direct readmission and increases the relative risk of readmission via ED (right panel in Fig S2-b). These results indicate the importance of accounting for the difference between discharge destinations, at least by separating "home with support" from "community or LTC" in future studies.

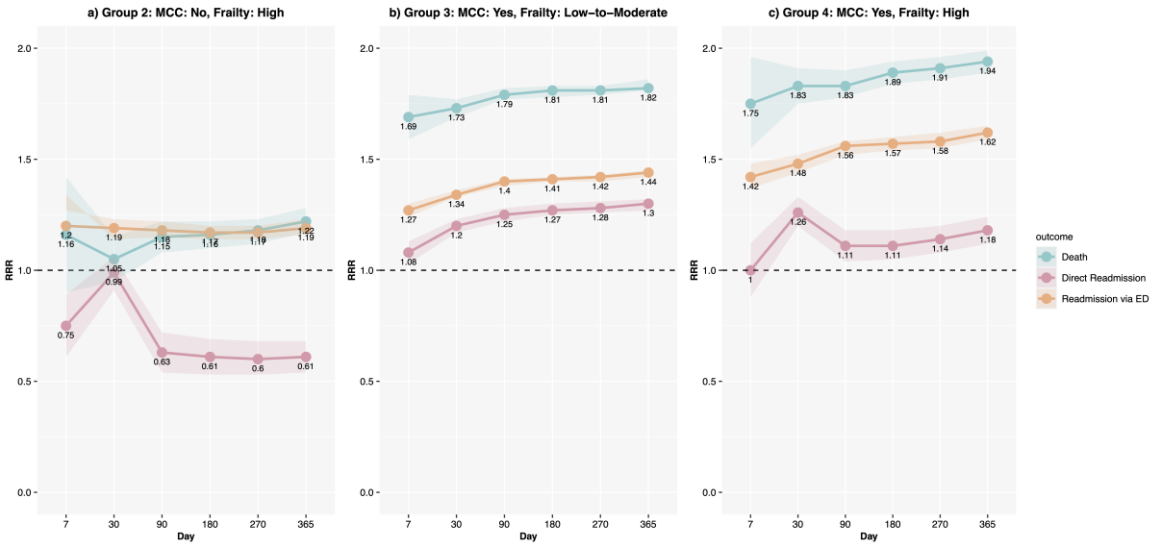


Fig. 4. The relative risk of outcomes in different time windows

Third, we investigated the impact of reconfiguring readmission outcome by i) combining both means of readmission and ii) excluding direct readmissions from the analysis. The results (Fig. S3 in Appendix A) suggest that these reconfigurations will bias the results, especially by overestimating the risk of direct readmission for all groups and underestimating the risk of ED readmission, especially for groups 2 and 4. These results reconfirm the CIHI's recommendation to separate these two means of hospital readmission, particularly among patients with discharge delay [60].

Fourth, we performed a sensitivity analysis to examine whether the results would differ between patients with short-moderate versus long wait times for discharge. To this end, we applied the 1.5×IQR (interquartile range) rule to identify the stratification [63]. Accordingly, we classified the points with wait times above "Q3+1.5×IQR" as potential long waiters, where Q3 corresponds to the third quartile. The results (Table S6 in Appendix B) suggest slight changes in the magnitude of RRR's but no change in their directions. Finally, we reconducted the primary analysis for

another cut-off point candidate (i.e., HFRS=5). Note that with this cut-off point, the dichotomized frailty variable must be interpreted as *low* vs. *moderate-high*. The results (Fig S4 in Appendix A) indicate that the associations are generally lower in magnitude but follow the same pattern. They also indicate a stronger discrimination power, in terms of patient outcomes, when using cut-off point = 14.

4. Discussion

To our knowledge, this is the first study to examine the impacts of multimorbidity and frailty on a series of patient-important outcomes, including mortality and hospital readmission among older adults with delayed discharge. The relative risks of mortality and readmission within 30 days were markedly higher in patients who had both MCC and a high level of frailty than the healthiest (or least complex) group (i.e., those without MCC who had a low-to-moderate level of frailty). Moreover, the effects of multimorbidity and frailty on the outcomes, especially mortality, were different for males versus females and varied across different window lengths, particularly after 90 days post-discharge.

4.1. Comparison of findings with prior medical literature

Our findings agree with previous studies providing evidence that simultaneously having MCC and a high level of frailty is associated with an increased risk of mortality and readmission in the community-dwelling older adults [37], general middle-aged and older adults [39], and patients with COVID-19 [38]. One study found that frailty was consistently associated with four-year mortality across various levels of multimorbidity after adjusting for age and sex [37]. The study reported that patients who were frail with multimorbidity (comparable to our group 4) had the

highest risk of four-year mortality compared to the least complex groups (comparable to our group 1). In a subgroup analysis examining the interaction between frailty and MCC count, Hanlon et al. [39] found that the likelihood of all-cause mortality among frail patients exponentially increased when the number of long-term comorbid conditions increased from one (odds ratio = 2.7) to at least four conditions (odds ratio = 27.1). The analyses were adjusted for chronic conditions, socioeconomic deprivation, and lifestyle factors (such as smoking and body mass index). Our findings are also consistent with a retrospective study reporting that frail patients with hypertension (comparable to our group 2) have a significantly higher risk of seven-day mortality among patients with COVID-19 [38]. Using large population-level data, our study suggests that frailty and multimorbidity are associated with an increased risk of mortality and readmission among older adults with delayed discharge.

Our results regarding the covariates-outcome association are also in agreement with the extant literature. For example, previous studies have demonstrated the association between marginalization, such as socioeconomic deprivation, and all-cause mortality [64,65] as well as 30-day readmission [66]. Moreover, the 30-day readmission and mortality rates were higher among rural residents [67]. Our findings regarding the greater risk of adverse events in male patients validate prior work highlighting higher rates of 30-day readmission [68,69] and mortality [70] in male patients. Finally, similar to our research, previous studies have reported that patients discharged to nursing homes [71] and LTC [72] were less likely to be readmitted within 30 days after discharge than those discharged home. Hoffman and colleagues concluded that hospital

readmissions were more likely in patients discharged home with support than in those with routine home discharge [73].

4.2. Health policy implications

This study highlights the importance of considering coexisting multimorbidity and frailty, in addition to several other patient-specific factors (such as sex, residency, and marginalization status), to better understand the complex needs of older delayed discharge patients and inform discharge policies by prioritizing patients at-risk for adverse outcomes. Advanced knowledge of these factors could support proactive, informed, and equitable discharge planning and clinical decision-making, given the greater risk of delayed discharge in older adults with complex conditions [26]. Clinicians and policymakers have advocated early discharge planning to identify the complex needs of geriatric patients and potential gaps in their transitions to community care [34,74,75], leading to better outcomes [76–78]. Embedding this information and triggering alert systems for high-risk patients within electronic medical charting may promote the feasibility and utility of these measures. In what follows, we discuss how this study provides insights into evidence-based policies for improving dispositions among older delayed discharge patients, some of which can also be applied to non-delayed discharge patients.

First, we demonstrated that multimorbidity and frailty each were strongly predictive of adverse outcomes. However, older adults with coexisting frailty and multimorbidity were at the greatest risk for mortality and hospital readmission. This finding validates the hypothesis that these measures provide unique information but are most informative when examined together. Therefore, vulnerability screeners and discharge planners for older hospitalized adults should

consider both measures simultaneously when stratifying patient risk in hospital settings. To this end, healthcare systems can use EHR data to develop screening tools to calculate informative measures of patients' health in terms of frailty and multimorbidity. This could be done more frequently to capture and track the changes in patients' complexity during hospitalization. Such information can, then, be used to categorize the delayed discharge patients into various risk groups, as suggested in this study, for managing their care, allocating the current resources (such as bed, staff, etc.), and planning the future resources based on the complexity of needs and risk profile of the patients in the hospital. This information can also be used for triggering alarms that facilitate early identification of high-risk patients and support decision-making regarding the patients' care pathway in multiple ways. First, it can complement and facilitate optimizing the use of advanced assessment approaches, such as Comprehensive Geriatric Assessment (CGA) tools [79]. Although these tools highly contribute to the patient-outcome improvements [79–81], their routine utilization for large cohorts of patients is challenging because of their high cost, workforce challenges, and time [82,83]. As suggested in this study, measuring patient complexity is less costly and more convenient, although it might be less accurate than CGA. The proposed approach in this study can identify priorities for using CGA to those of more complex needs, hence optimizing the CGA utilization based on the available resources. Second, detecting the delayed discharge patients at risk of adverse outcomes can help discharge planners identify appropriate care pathways post-discharge. Some of the delayed discharge patients end up going home (with or without support) instead of more intensive care settings (such as LTC or rehabilitation) due to financial reasons on the patient side or limited resources on the system side. However, there is evidence that such a pathway is temporary [84] and, as our study suggests, sub-optimal (because

of their higher post-discharge mortality and rehospitalization). Therefore, a routine assessment of the patient complexity provides valuable information at the right time regarding the right care pathway for the patients and the right intensity of services that they need to be discharged home safely and maintain health and functioning during their temporary settlements at home.

Second, there is evidence that singular interventions are ineffective in significantly improving patient outcomes (such as death and rehospitalization) [85]. Moreover, despite the benefits of post-discharge interventions, such as follow-up visits by an outgoing multidisciplinary geriatric team, they are costly and may not be applicable to a large cohort of patients [86]. Therefore, aligning interventions with the desired patient outcomes through the regular review and evaluations of the patients at-risk for post-discharge adverse outcomes can be beneficial. In this regard, our findings can inform tailoring these strategies to the patient complexity or prioritizing patients with a higher risk of adverse outcomes.

Third, our study demonstrated that the impact of multimorbidity and frailty on the adverse outcomes depends on biological sex, highlighting this variable as an important consideration when determining those who may benefit from immediate MCC-Frailty risk screening following hospital admission. Therefore, examining sex is essential in aging research, as health service use and the physiological and psychosocial aging process are vastly different between male and female patients [87,88]. Moreover, our findings suggest that older patients from rural areas or marginalized groups have a greater risk of adverse health outcomes when facing delayed discharge. Patients living in rural areas are known to have fewer health services available and consequently worse patient outcomes [89], emphasizing the importance of considering the location

of permanent dwellings. This is particularly important in the delayed discharge patients residing in rural areas, as they will be returning to the community with fewer resources in a deconditioned and vulnerable state. In this regard, health equity reinforcements have been identified as a policy priority by the Ontario Ministry of Health and Ontario Ministry of Long-term Care, encouraging providers and policymakers to include equity attentions to reduce disparities between marginalized groups, such as women and racialized and immigrant groups [90]. Our study further highlights such equity-seeking considerations, particularly among patients experiencing discharge delay, to adapt care practices and policies that promote equity-oriented care to improve patient outcomes.

Finally, a common priority-based discharge policy regarding access to community care is the first-come-first-served (FCFS) policy that prioritizes patients according to their wait time for access [33,91,92]. However, evidence suggests that tailoring hospital and health policies to account for the patient complexity and needs is better aligned with the philosophy of patient-centered care [34,93–96] and can lead to better outcomes [33]. Our study provides empirical evidence regarding this recommendation (Fig. 2). Our results suggest that an increased wait time for the alternative care placement (i.e., increased ALCW) leads to small changes in the adverse outcomes compared to the significant increase in the risk of adverse outcomes caused by aggravation in the MCC-Frailty status. Future studies could investigate the implications of increased wait time for designing risk-adjusted, equity-seeking FCFS policies.

We speculate that our study's insights about the pre- and post-discharge policies would still apply to the non-delayed discharge older adults. For instance, regardless of the delayed discharge designations, patient complexity could be assessed and reported to the clinical and managerial

teams during hospitalization to inform the intensity and type of post-discharge care for the patients (even those discharged without delay).

Moving toward utilizing succinct and informative indices, such as ECI and HFRS, using EHR, might be advantageous for a rapid and routine assessment of frailty and MCC [42,97,98]. However, we note that there is always a trade-off between convenience and data quality when using EHR data [99].

4.3. Limitations

Our study is not without limitations. First, we used a cut-off score of HFRS = 14 as the best dichotomization threshold. The original article [42] has suggested 5 and 15 as the two thresholds for defining low, moderate, and high frailty. Our numerical approach (Appendix C) includes investigating these two cut-off points and sensitivity analysis around them to select the one that best discriminates among patients with and without the outcomes. Second, this is a retrospective cohort study with limited control over data collection and validation; however, data quality control is performed regularly by the IC/ES organization. Moreover, frailty is best measured prospectively; however, like prior studies, we applied the calculations retrospectively given the nature of data collection in our study. Fourth, upcoming studies are encouraged to investigate the sensitivity of the results to the acute patients who are likely to be designated as delayed discharge but underreported in the database. This could be done using the Case Mix Groups+ approach suggested by the CIHI [100]. Fifth, further research is warranted to extend our analyses to the general population or compare older adults with and without discharge delay. Finally, there might have been other confounders that were not included in our data.

5. Conclusion

To our knowledge, this is the first study of patients with delayed discharge that investigates the association of frailty and comorbidity with post-discharge mortality and readmission. It provides actionable insights into evidence-based policies for improving dispositions among older delayed discharge patients. Using large data, our study suggests that multimorbidity and high frailty markedly increase the risk of adverse outcomes. Moreover, being male or living in a rural/remote area increases the risk of readmission (both via ED and direct) as well as mortality within 30 days after discharge. Higher marginalization will increase the risk of mortality and readmission via ED but decrease the risk of direct readmission. Our study emphasizes the importance of routine assessments of multimorbidity and frailty to improve risk prediction and facilitate individualized care management for the aging population, particularly those with delayed discharge.

References

[1] Vetrano DL, Palmer K, Marengoni A, Marzetti E, Lattanzio F, Roller-Wirnsberger R, et al. Frailty and Multimorbidity: A Systematic Review and Meta-analysis. J Gerontol A Biol Sci Med Sci 2019;74:659–66. https://doi.org/10.1093/gerona/gly110.

[2] Maciejewski ML, Hammill BG. Measuring the burden of multimorbidity among Medicare beneficiaries via condition counts and cumulative duration. Health Serv Res 2019;54:484–91.

[3] Jurevičienė E, Onder G, Visockienė Ž, Puronaitė R, Petrikonytė D, Gargalskaitė U, et al. Does multimorbidity still remain a matter of the elderly: Lithuanian national data analysis. Health Policy 2018;122:681–6. https://doi.org/10.1016/j.healthpol.2018.03.003.

[4] Rijken M, Hujala A, van Ginneken E, Melchiorre MG, Groenewegen P, Schellevis F. Managing multimorbidity: Profiles of integrated care approaches targeting people with multiple chronic conditions in Europe. Health Policy 2018;122:44–52. https://doi.org/10.1016/j.healthpol.2017.10.002.

[5] Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56:M146-156. https://doi.org/10.1093/gerona/56.3.m146.

[6] Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ Can Med Assoc J J Assoc Medicale Can 2005;173:489–95. https://doi.org/10.1503/cmaj.050051.

[7] Han L, Clegg A, Doran T, Fraser L. The impact of frailty on healthcare resource use: a longitudinal analysis using the Clinical Practice Research Datalink in England. Age Ageing 2019;48:665–71. https://doi.org/10.1093/ageing/afz088.

[8] Mowbray F, Brousseau A-A, Mercier E, Melady D, Émond M, Costa AP. Examining the relationship between triage acuity and frailty to inform the care of older emergency department patients: Findings from a large Canadian multisite cohort study. Can J Emerg Med 2020;22:74–81. https://doi.org/10.1017/cem.2019.432.

[9] Caballer-Tarazona V, Guadalajara-Olmeda N, Vivas-Consuelo D. Predicting healthcare expenditure by multimorbidity groups. Health Policy 2019;123:427–34.

[10] Schaink AK, Kuluski K, Lyons RF, Fortin M, Jadad AR, Upshur R, et al. A scoping review and thematic classification of patient complexity: offering a unifying framework. J Comorbidity 2012;2:1–9. https://doi.org/10.15256/joc.2012.2.15.

[11] Nardi R, Scanelli G, Corrao S, Iori I, Mathieu G, Amatrian RC. Co-morbidity does not reflect complexity in internal medicine patients. Eur J Intern Med 2007;18:359–68.

[12] Lacas A, Rockwood K. Frailty in primary care: a review of its conceptualization and implications for practice. BMC Med 2012;10:4. https://doi.org/10.1186/1741-7015-10-4.

[13] McGeorge SJ. Unravelling the differences between complexity and frailty in old age: findings from a constructivist grounded theory study. J Psychiatr Ment Health Nurs 2011;18:67–73. https://doi.org/10.1111/j.1365-2850.2010.01630.x.

[14] Creditor MC. Hazards of hospitalization of the elderly. Ann Intern Med 1993;118:219–23. https://doi.org/10.7326/0003-4819-118-3-199302010-00011.

[15] Covinsky KE, Palmer RM, Fortinsky RH, Counsell SR, Stewart AL, Kresevic D, et al. Loss of independence in activities of daily living in older adults hospitalized with medical illnesses: increased vulnerability with age. J Am Geriatr Soc 2003;51:451–8. https://doi.org/10.1046/j.1532-5415.2003.51152.x.

[16] Rothschild JM, Bates DW, Leape LL. Preventable medical injuries in older patients. Arch Intern Med 2000;160:2717–28. https://doi.org/10.1001/archinte.160.18.2717.

[17] Rosenthal GE, Kaboli PJ, Barnett MJ, Sirio CA. Age and the risk of in-hospital death: insights from a multihospital study of intensive care patients. J Am Geriatr Soc 2002;50:1205–12. https://doi.org/10.1046/j.1532-5415.2002.50306.x.

[18] Barnett ML, Hsu J, McWilliams JM. Patient Characteristics and Differences in Hospital Readmission Rates. JAMA Intern Med 2015;175:1803–12. https://doi.org/10.1001/jamainternmed.2015.4660.
[19] Bottle A, Jarman B, Aylin P. Strengths and weaknesses of hospital standardised mortality ratios. BMJ 2010;342:c7116. https://doi.org/10.1136/bmj.c7116.
[20] Fischer C, Lingsma HF, Marang-van de Mheen PJ, Kringos DS, Klazinga NS, Steyerberg EW. Is the Readmission Rate a Valid Quality Indicator? A Review of the Evidence. PLoS ONE 2014;9. https://doi.org/10.1371/journal.pone.0112282.
[21] Akpan A, Roberts C, Bandeen-Roche K, Batty B, Bausewein C, Bell D, et al. Standard set of health outcome measures for older persons. BMC Geriatr 2018;18:36. https://doi.org/10.1186/s12877-017-0701-3.
[22] McGilton KS, Vellani S, Babineau J, Bethell J, Bronskill SE, Burr E, et al. Understanding transitional care programmes for older adults who experience delayed discharge: a scoping review protocol. BMJ Open 2019;9.
[23] Manzano-Santaella A. From bed-blocking to delayed discharges: precursors and interpretations of a contested concept. Health Serv Manage Res 2010;23:121–7.
[24] Challis D, Hughes J, Xie C, Jolley D. An examination of factors influencing delayed discharge of older people from hospital. Int J Geriatr Psychiatry 2014;29:160–8.
[25] McGilton KS, Vellani S, Babineau J, Bethell J, Bronskill SE, Burr E, et al. Understanding transitional care programmes for older adults who experience delayed discharge: a scoping review protocol. BMJ Open 2019;9.
[26] Bhatia D, Peckham A, Abdelhalim R, King M, Kurdina A, Ng R, et al. Alternate Level of Care and Delayed Discharge: Lessons Learned from Abroad. Toronto: North American Observatory on Health Systems and Policies. Rapid Review (No. 22); 2020.
[27] Gaughan J, Gravelle H, Siciliani L. Delayed discharges and hospital type: evidence from the English NHS. Fisc Stud 2017;38:495–519.
[28] Landeiro F, Roberts K, Gray AM, Leal J. Delayed hospital discharges of older patients: a systematic review on prevalence and costs. The Gerontologist 2019;59:e86–97.
[29] Zychlinski N, Mandelbaum A, Momčilović P, Cohen I. Bed Blocking in Hospitals Due to Scarce Capacity in Geriatric Institutions—Cost Minimization via Fluid Models. Manuf Serv Oper Manag 2019;22:396–411. https://doi.org/10.1287/msom.2018.0745.
[30] Rojas-García A, Turner S, Pizzo E, Hudson E, Thomas J, Raine R. Impact and experiences of delayed discharge: A mixed-studies systematic review. Health Expect Int J Public Particip Health Care Health Policy 2018;21:41–56. https://doi.org/10.1111/hex.12619.
[31] Landeiro F, Roberts K, Gray AM, Leal J. Delayed Hospital Discharges of Older Patients: A Systematic Review on Prevalence and Costs. The Gerontologist 2019;59:e86–97. https://doi.org/10.1093/geront/gnx028.
[32] McCloskey R, Jarrett P, Stewart C, Nicholson P. Alternate level of care patients in hospitals: what does dementia have to do with this? Can Geriatr J 2014;17:88.
[33] Chafe R, Coyte P, Sears NA. Improving the management of waiting lists for long term care. Healthc. Manage. Forum, vol. 23, SAGE Publications Sage CA: Los Angeles, CA; 2010, p. 58–62.
[34] Bakerjian D, Vanraay E, Ferris B. Placement, Coordination, and Follow-Up. In: Fenn HH, Hategan A, Bourgeois JA, editors. Inpatient Geriatr. Psychiatry Optim. Care Emerg. Limit. Realistic Goals, Cham: Springer International Publishing; 2019, p. 393–404. https://doi.org/10.1007/978-3-030-10401-6_21.
[35] Amy C, Zagorski B, Chan V, Parsons D, Vander Laan R, Colantonio A. Acute Care Alternate-Level-of-Care Days Due to Delayed Discharge for Traumatic and Non-Traumatic Brain Injuries. Healthc Policy 2012;7:41–55.
[36] Bender D, Holyoke P. Why some patients who do not need hospitalization cannot leave: A case study of reviews in 6 Canadian hospitals. Healthc Manage Forum 2018;31:121–5. https://doi.org/10.1177/0840470418755408.

[37] Nguyen QD, Wu C, Odden MC, Kim DH. Multimorbidity patterns, frailty, and survival in community-dwelling older adults. J Gerontol Ser A 2019;74:1265–70.
[38] Hewitt J, Carter B, Vilches-Moraga A, Quinn TJ, Braude P, Verduri A, et al. The effect of frailty on survival in patients with COVID-19 (COPE): a multicentre, European, observational cohort study. Lancet Public Health 2020;5:e444–51. https://doi.org/10.1016/S2468-2667(20)30146-8.
[39] Hanlon P, Nicholl BI, Jani BD, Lee D, McQueenie R, Mair FS. Frailty and pre-frailty in middle-aged and older adults and its association with multimorbidity and mortality: a prospective analysis of 493 737 UK Biobank participants. Lancet Public Health 2018;3:e323–32.
[40] Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998;36:8–27. https://doi.org/10.1097/00005650-199801000-00004.
[41] Azzalini L, Chabot-Blanchet M, Southern DA, Nozza A, Wilton SB, Graham MM, et al. A disease-specific comorbidity index for predicting mortality in patients admitted to hospital with a cardiac condition. CMAJ 2019;191:E299–307.
[42] Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. The Lancet 2018;391:1775–82.
[43] Kim LH, Chen Y-R. Risk Adjustment Instruments in Administrative Data Studies: A Primer for Neurosurgeons. World Neurosurg 2019;128:477–500.
[44] Qiukui H, Zhou L, Biao D, Yang M, Birong D, Yuquan W. The role of frailty in predicting mortality and readmission in older adults in acute care wards: a prospective study. Sci Rep Nat Publ Group 2019;9.
[45] Qureshi D, Tanuseputro P, Perez R, Seow H. Place of Care Trajectories in the Last Two Weeks of Life: A Population-Based Cohort Study of Ontario Decedents. J Palliat Med 2018;21:1588–95. https://doi.org/10.1089/jpm.2018.0049.
[46] Ryan B, Allen B, Zwarenstein M, Stewart M, Glazier R, Fortin M, et al. Multimorbidity and mortality in Ontario, Canada: A population-based retrospective cohort study. J Comorbidity 2020;10:2235042X20950598. https://doi.org/10.1177/2235042X20950598.
[47] Jones A, Bronskill SE, Seow H, Junek M, Feeny D, Costa AP. Associations between continuity of primary and specialty physician care and use of hospital-based care among community-dwelling older adults with complex care needs. PloS One 2020;15:e0234205.
[48] McAlister F, van Walraven C. External validation of the Hospital Frailty Risk Score and comparison with the Hospital-patient One-year Mortality Risk Score to predict outcomes in elderly hospitalised patients: a retrospective cohort study. BMJ Qual Saf 2019;28:284–8.
[49] Canadian Institute for Health Information. Definitions and Guidelines to Support ALC Designation in Acute Inpatient Care. 2016.
[50] Centers for Medicare & Medicaid Services. 2015 measure information about the 30-day all-cause hospital readmission measure, calculated for the value-based payment modifier program (Updated March 2017). 2017.
[51] Shahian DM, Liu X, Mort EA, Normand S-LT. The association of hospital teaching intensity with 30-day postdischarge heart failure readmission and mortality rates. Health Serv Res 2020;55:259–72. https://doi.org/10.1111/1475-6773.13248.
[52] Fauchier L, Bisson A, Bodin A, Clementy N, Pierre B, Angoulvant D, et al. P4748 HASBLED score, frailty index or comorbidity index for bleeding risk assessment in patients with atrial fibrillation? Eur Heart J 2019;40:ehz745-1124.
[53] van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. J Clin Epidemiol 2004;57:672–82.
[54] Senot C. Continuity of care and risk of readmission: An investigation into the healthcare journey of heart failure patients. Prod Oper Manag 2019;28:2008–30.
[55] Matheson FI, van Ingen. 2016 Ontario marginalization index. Tor St Michael's Hosp 2018 Jt Publ Public Health Ont 2018.

[56] Halwani S. Racial inequality in access to health care services. Pap Ont Hum Rights Comm 2002.
[57] Reitz JG, Banerjee R. Racial inequality, social cohesion and policy issues in Canada. Citeseer; 2007.
[58] Canadian Institute for Health Information. In Pursuit of Health Equity: Defining Stratifiers for Measuring Health Inequality — A Focus on Age, Sex, Gender, Income, Education and Geographic Location 2018.
[59] Mowbray F, Zargoush M, Jones A, de Wit K, Costa A. Predicting hospital admission for older emergency department patients: Insights from machine learning. Int J Med Inf 2020;140:1–8. https://doi.org/10.1016/j.ijmedinf.2020.104163.
[60] Dawson H, Weerasooriya J, Webster G. Hospital admissions via the emergency department: implications for planning and patient flow. Healthc Q Tor Ont 2008;11:20–2. https://doi.org/10.12927/hcq.2013.19492.
[61] Leyenaar JK, Lagu T, Lindenauer PK. Direct admission to the hospital: an alternative approach to hospitalization. J Hosp Med 2016;11:303–5.
[62] Powell ES, Khare RK, Courtney DM, Feinglass J. Lower mortality in sepsis patients admitted through the ED vs direct admission. Am J Emerg Med 2012;30:432–9.
[63] Hubert M, Vandervieren E. An adjusted boxplot for skewed distributions. Comput Stat Data Anal 2008;52:5186–201.
[64] Foster HME, Celis-Morales CA, Nicholl BI, Petermann-Rocha F, Pell JP, Gill JMR, et al. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. Lancet Public Health 2018;3:e576–85. https://doi.org/10.1016/S2468-2667(18)30200-7.
[65] Zygmunt A, Tanuseputro P, James P, Lima I, Tuna M, Kendall CE. Neighbourhood-level marginalization and avoidable mortality in Ontario, Canada: a population-based study. Can J Public Health 2019:1–13.
[66] Galiatsatos P, Follin A, Alghanim F, Sherry M, Sylvester C, Daniel Y, et al. The Association Between Neighborhood Socioeconomic Disadvantage and Readmissions for Patients Hospitalized With Sepsis. Crit Care Med 2020;48:808–14. https://doi.org/10.1097/CCM.0000000000004307.
[67] Kosar CM, Loomer L, Ferdows NB, Trivedi AN, Panagiotou OA, Rahman M. Assessment of Rural-Urban Differences in Postacute Care Utilization and Outcomes Among Older US Adults. JAMA Netw Open 2020;3:e1918738–e1918738.
[68] Tan SY, Feng JY, Joyce C, Fisher J, Mostaghimi A. Association of Hospital Discharge Against Medical Advice With Readmission and In-Hospital Mortality. JAMA Netw Open 2020;3:e206009–e206009.
[69] Buja A, De Polo A, De Battisti E, Sperotto M, Baldovin T, Cocchio S, et al. The importance of sex as a risk factor for hospital readmissions due to pulmonary diseases. BMC Public Health 2020;20:53. https://doi.org/10.1186/s12889-019-8138-6.
[70] Hakkarainen TW, Arbabi S, Willis MM, Davidson GH, Flum DR. Outcomes of patients discharged to skilled nursing facilities after acute care hospitalizations. Ann Surg 2016;263:280.
[71] Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med 2000;160:1074–81.
[72] Gruneir A, Fung K, Fischer HD, Bronskill SE, Panjwani D, Bell CM, et al. Care setting and 30-day hospital readmissions among older adults: a population-based cohort study. CMAJ 2018;190:E1124–33.
[73] Hoffman GJ, Liu H, Alexander NB, Tinetti M, Braun TM, Min LC. Posthospital fall injuries and 30-day readmissions in adults 65 years and older. JAMA Netw Open 2019;2:e194276–e194276.
[74] Ohta B, Mola A, Rosenfeld P, Ford S. Early discharge planning and improved care transitions: pre-admission assessment for readmission risk in an elective orthopedic and cardiovascular surgical population. Int J Integr Care 2016;16.
[75] Coffey A, Leahy-Warren P, Savage E, Hegarty J, Cornally N, Day MR, et al. Interventions to Promote Early Discharge and Avoid Inappropriate Hospital (Re)Admission: A Systematic Review. Int J Environ Res Public Health 2019;16. https://doi.org/10.3390/ijerph16142457.
[76] Fox MT, Persaud M, Maimets I, Brooks D, O'Brien K, Tregunno D. Effectiveness of early discharge planning in acutely ill or injured hospitalized older adults: a systematic review and meta-analysis. BMC Geriatr 2013;13:70. https://doi.org/10.1186/1471-2318-13-70.

[77] Burke RE, Guo R, Prochazka AV, Misky GJ. Identifying keys to success in reducing readmissions using the ideal transitions in care framework. BMC Health Serv Res 2014;14:423. https://doi.org/10.1186/1472-6963-14-423.
[78] Dyrstad DN, Testad I, Aase K, Storm M. A review of the literature on patient participation in transitions of the elderly. Cogn Technol Work 2015;17:15–34. https://doi.org/10.1007/s10111-014-0300-4.
[79] Ellis G, Gardner M, Tsiachristas A, Langhorne P, Burke O, Harwood RH, et al. Comprehensive geriatric assessment for older adults admitted to hospital. Cochrane Database Syst Rev 2017.
[80] Pickert L, Meyer AM, Becker I, Heeß A, Noetzel N, Brinkkötter P, et al. Role of a multidimensional prognosis in-hospital monitoring for older patients with prolonged stay. Int J Clin Pract 2021;75:e13989. https://doi.org/10.1111/ijcp.13989.
[81] Shepperd S, Butler C, Cradduck-Bamford A, Ellis G, Gray A, Hemsley A, et al. Is Comprehensive Geriatric Assessment Admission Avoidance Hospital at Home an Alternative to Hospital Admission for Older Persons? Ann Intern Med 2021;174:889–98. https://doi.org/10.7326/M20-5688.
[82] Gladman JR, Conroy SP, Ranhoff AH, Gordon AL. New horizons in the implementation and research of comprehensive geriatric assessment: knowing, doing and the 'know-do'gap. Age Ageing 2016;45:194–200.
[83] Palumbo A, Bringhen S, Mateos M-V, Larocca A, Facon T, Kumar SK, et al. Geriatric assessment predicts survival and toxicities in elderly myeloma patients: an International Myeloma Working Group report. Blood J Am Soc Hematol 2015;125:2068–74.
[84] Walker D, Lead PA. Caring for our aging population and addressing alternate level of care. Ontario Ministry of Health and Long-Term Care; 2011.
[85] Kripalani S, Theobald CN, Anctil B, Vasilevskis EE. Reducing hospital readmission rates: current strategies and future directions. Annu Rev Med 2014;65:471–85.
[86] Thomsen K, Fournaise A, Matzen LE, Andersen-Ranberg K, Ryg J. Does geriatric follow-up visits reduce hospital readmission among older patients discharged to temporary care at a skilled nursing facility: a before-and-after cohort study. BMJ Open 2021;11:e046698.
[87] Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. BMC Fam Pract 2016;17:38. https://doi.org/10.1186/s12875-016-0440-0.
[88] Rosenfield S, Mouzon D. Gender and mental health. Handb. Sociol. Ment. Health, Springer; 2013, p. 277–96.
[89] DesMeules M, Pong R, Lagacé C, Heng D, Manuel D, Pitblado R, et al. How healthy are rural Canadians? An assessment of their health status and health determinants. Can Inst Health Inf 2006.
[90] Greenwood M, de Leeuw S, Lindsay N. Challenges in health equity for Indigenous peoples in Canada. The Lancet 2018;391:1645–8.
[91] Noghani Ardestani P. Modeling Community Care Services for Alternative level of Care (ALC) Patients: A Queuing Network Approach. PhD Thesis. Université d'Ottawa/University of Ottawa, 2014.
[92] Meadus JE. Discharge from hospital to long-term care: issues in Ontario. Advocacy Cent. Elder. Newsl., 2010.
[93] Zargoush M, Gümüş M, Verter V, Daskalopoulou SS. Designing Risk-Adjusted Therapy for Patients with Hypertension. Prod Oper Manag 2018;27:2291–312. https://doi.org/10.1111/poms.12872.
[94] Reynolds A. Patient-centered care. Radiol Technol 2009;81:133–47.
[95] Kuipers SJ, Nieboer AP, Cramm JM. Making care more patient centered; experiences of healthcare professionals and patients with multimorbidity in the primary care setting. BMC Fam Pract 2021;22:1–15.
[96] van der Heide I, Snoeijs S, Quattrini S, Struckmann V, Hujala A, Schellevis F, et al. Patient-centeredness of integrated care programs for people with multimorbidity. Results from the European ICARE4EU project. Health Policy 2018;122:36–43. https://doi.org/10.1016/j.healthpol.2017.10.005.

[97] Hanlon P, Fauré I, Corcoran N, Butterly E, Lewsey J, McAllister D, et al. Frailty measurement, prevalence, incidence, and clinical implications in people with diabetes: a systematic review and study-level meta-analysis. Lancet Healthy Longev 2020.
[98] Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. The Lancet 2019;394:1365–75.
[99] Timofte D, Pantea Stoian A, Razvan H. A review on the advantages and disadvantages of using administrative data in surgery outcome studies. J Surg Jurnalul Chir 2018;14:105–7.
[100] Canadian Institute for Health Informatio. Discharge Abstract Database Open-Year Data Quality Test Specifications, 2020–2021 2021.

Supplementary Appendix

TABLE OF CONTENTS

APPENDIX A LIST OF FIGURES

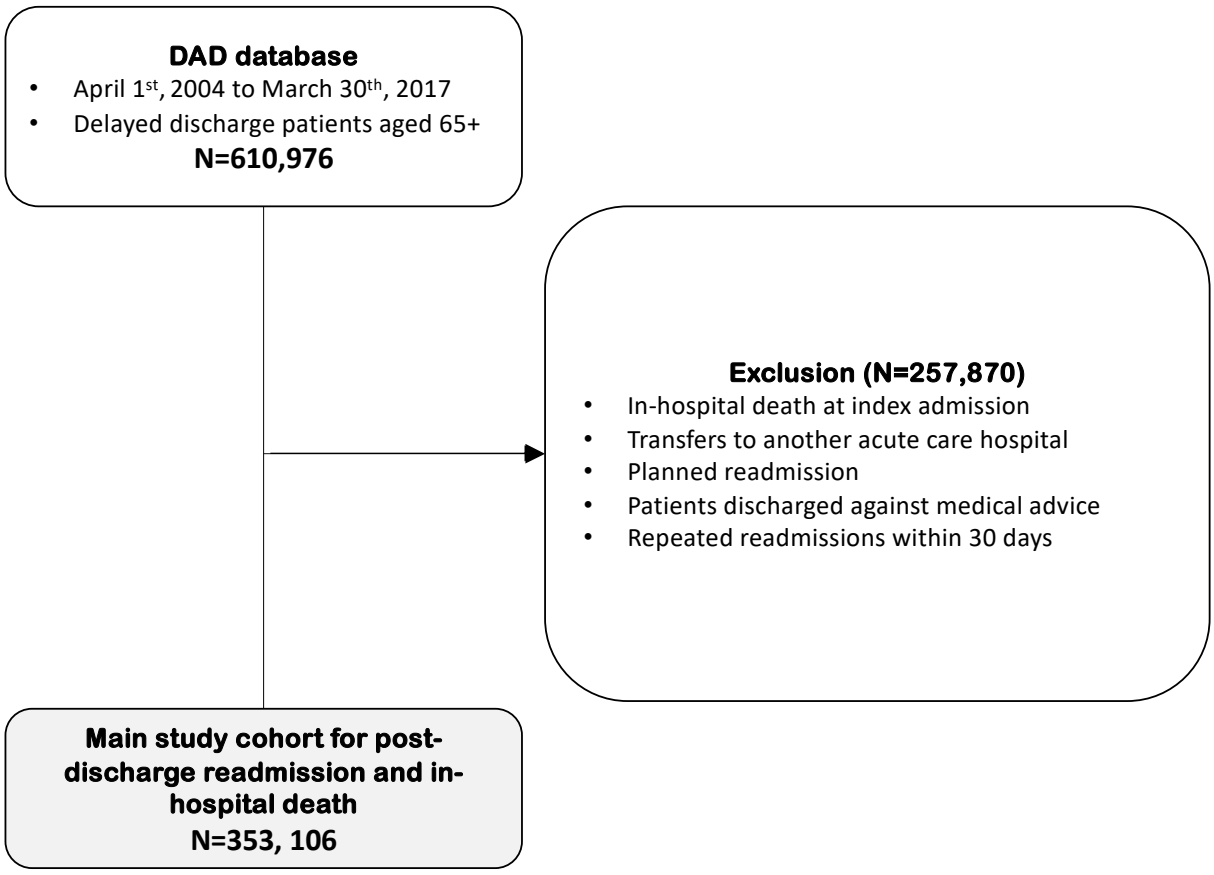


Fig. S1. Flowchart of patient exclusions

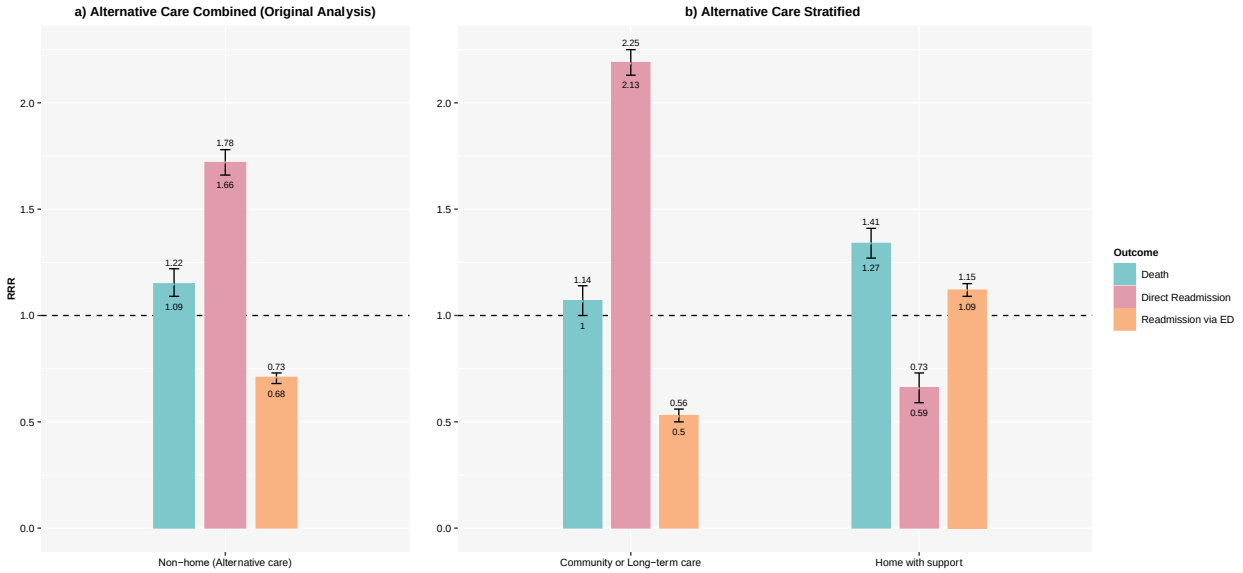


Fig. S2. Comparing outcomes across discharge destinations (ref. is home with self-care)

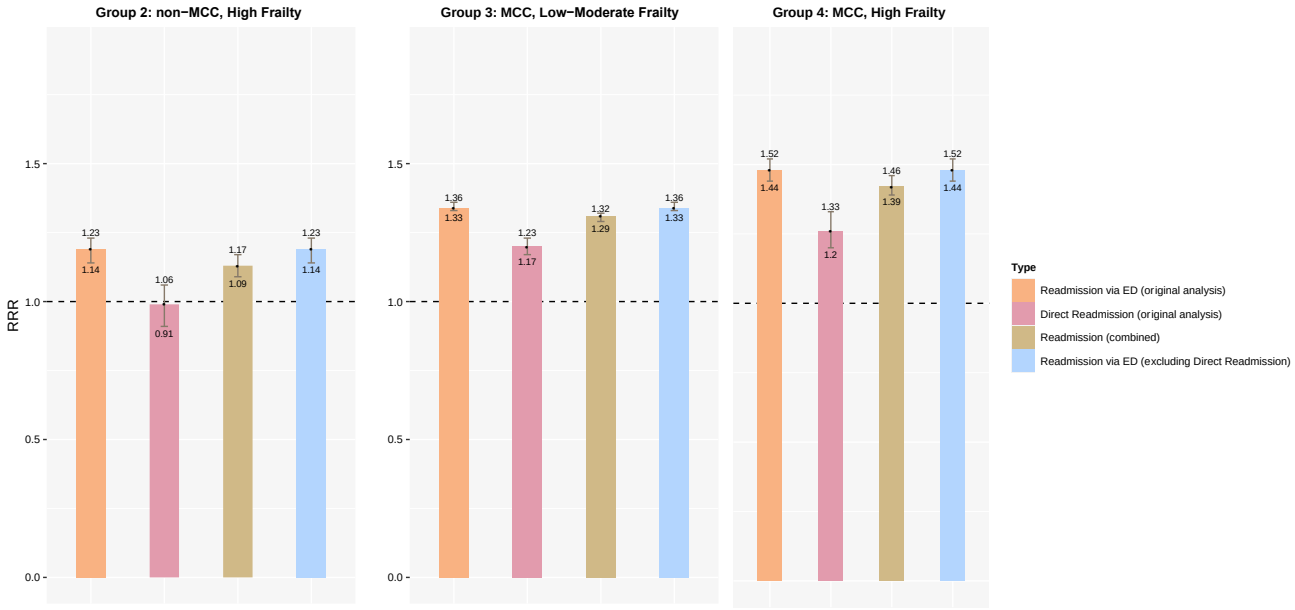


Fig. S3. Comparing risks for various configurations of readmission

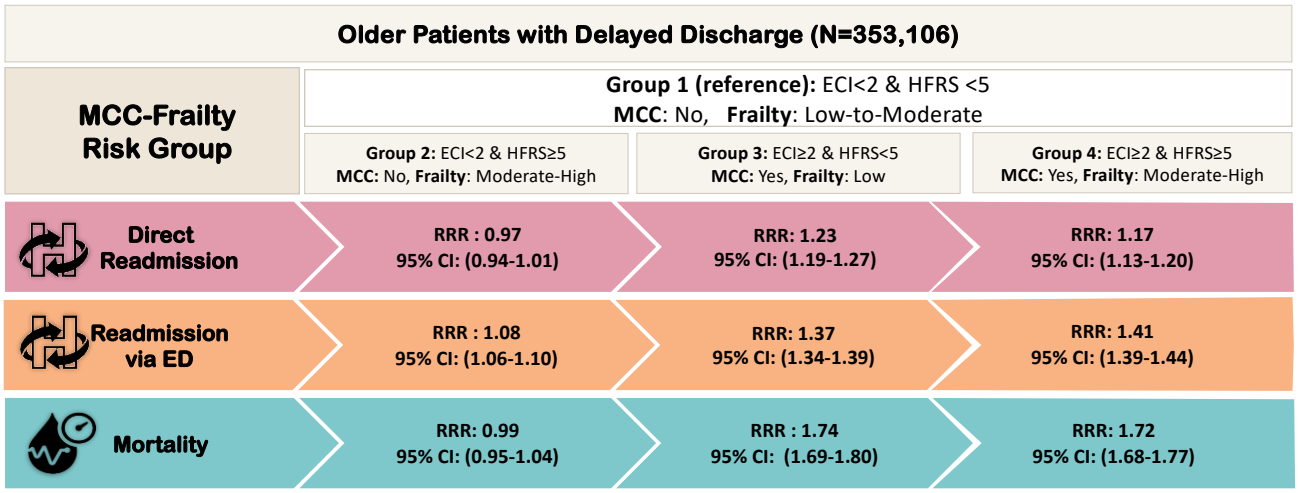


Fig. S4. Adjusted RRR for the MCC-Frailty risk categories with candidate cutoff point=5 for HFRS
Note: HFRS: hospital frailty risk score, ECI: Elixhauser Comorbidity Index, RRR: relative risk ratio

APPENDIX B LIST OF TABLES

Table S1. Literature Review Summary

Study	Year	Country	#Obs.	Setting	Design	Outcome	Frailty measure	Comorbidity measure	Covariates	analysis	MCC/ Frailty coexistence?
Kahlon et al. [1]	2015	Canada	495	general internal medicine wards (>18yrs)	Prospective	30-day readmission or death	CFS	CCI applied in LACE score	age, sex, LACE	adjustment for comorbidity	No
Qiukui et al. [2]	2019	China	271	acute care wards (older adults)	Prospective	3-year mortality and readmission	FI	Comorbidity-D-FI	age, sex, education levels, BMI, marital status, and alcohol intake	adjustment for comorbidity	No
Li et al. [3]	2018	Canada	322	patients undergoing emergency abdominal surgery (>65yrs)	Prospective	30-day and 6-month all-cause readmission or death	CFS	CCI	age, sex, type of surgery, some biological markers	adjustment for comorbidity	No
Vidan et al. [4]	2016	Italy	450	non-dependent Heart Failure patients (>70yrs)	Prospective	30-day functional decline, 1-year all-cause mortality and readmission	phenotype model	CCI	(age, gender), chronic co-morbidity, and some biologic markers	adjustment for comorbidity	No
Tran et al. [5]	2018	Canada	40,083	patients undergoing CABG (>40yrs)	Retrospective	30-day and long-term mortality	ACG	multiple specific chronic conditions	(age, sex), socioeconomic status, case urgency status, comorbidities	adjustment for comorbidity	No
Hatcher et al. [6]	2019	USA	804	patients with a trauma-related injury who readmitted for fall (>50yrs)	Retrospective	mortality rate, discharge disposition, LOS, and number of falls within 1-year	CFC	multiple specific chronic conditions	age, BMI, sex, anticoagulant or aspirin use, and some clinical markers related to injury patterns (SSI)	adjustment for comorbidity	No
Hewitt et al.[7]	2020	UK & Italy	1564	patients with COVID19 (>18yrs)	Retrospective	seven-day mortality, in-hospital mortality, and LOS	CFS	multiple specific chronic conditions	age, comorbidity, gender, and smoking	adjustment for comorbidity, subgroup analysis for testing the interactive effects	Yes

Study	Year	Country	#Obs.	Setting	Design	Outcome	Frailty measure	Comorbidity measure	Covariates	analysis	MCC/ Frailty coexistence?
Nguyen et al. [8]	2019	USA	7,197	community-dwelling older adults (>65yrs)	Prospective	four-year mortality incidence rate	phenotype model	five multimorbidity patterns	age and sex	investigating the change in the impact of frailty on the outcomes through different multimorbidity patterns	Yes
Ritt et al. [9]	2017	Germany	307	geriatric wards (>65yrs)	Prospective	one-year mortality	CFS, FI, FRAIL	CIRS-G and Comorbidity-D-FI	age and gender	compare the prognostic effect of frailty, comorbidity, and disability,	No
Hanlon et al. [10]	2018	UK	493,737	UK bionak (37-73yrs)	Prospective	seven-year all-cause mortality	phenotype model	five groups of multimorbidity categorized by count (0 to 4+)	age and sex, multimorbidity count, socioeconomic deprivation, body-mass index, smoking status, and alcohol use	adjustment for comorbidity, subgroup analysis for testing the interactive effects	Yes
Serina et al. [11]	2020	USA	7,304	ED patients (>65yrs)	Retrospective	Hospital admission at the time of ED visit; ED return visit within 9 days; readmission within 30 days of ED visit	CFS	-	No confounder	Unadjusted model for determining the impact of frailty on the outcome	No

Table S2. Patient characteristics in terms of stratified groups

Risk Groups	Group 1 (ECI<2, HFRS≤14)	Group 2 (ECI<2, HFRS>14)	Group 3 (ECI≥2, HFRS≤14)	Group 4 (ECI≥2, HFRS>14)	Overall
Characteristics	219,057 (62%)	13,497 (3.8%)	105,321 (29.8%)	15,231 (4.3%)	353,106 (100%)
Age, mean (SD)	79.1 (7.9)	80.4 (7.4)	76.8 (7.9)	77.5 (7.7)	78.3 (7.9)
Sex					
Male	84,989 (39%)	5,240 (39%)	47,792 (45%)	6,701 (44%)	144,722 (41%)
Female	134,068 (61%)	8,257 (61%)	57,529 (55%)	8,530 (56%)	208,384 (59%)
Residency					
Rural	27,316 (12%)	889 (6.6%)	10,815 (10%)	928 (7%)	39,948 (11.4%)
Urban	191,741 (88%)	12,608(93.4%)	94,506 (90%)	14,303 (93%)	313,158 (88.6%)
Discharge destination					
home	20,589 (9%)	608 (4.5 %)	10,549 (10%)	703 (4.6%)	32,449 (9%)
Non-home	198,468 (91%)	12,889 (95.5%)	94,772 (90%)	14,528 (95.4%)	320,657 (91%)
Material deprivation					
Quintile 1(least)	32,613 (15%)	2,380 (17.6%)	14,782 (14.2%)	2,388 (15.7%)	52,163 (14.8%)
Quintile 2	37,438 (17.3%)	2,342(17.4%)	17,140 (16.3%)	2,482 (16.3%)	59,402 (16.8%)
Quintile 3	42,634 (19.4%)	2,421 (17.9%)	19,679 (18.7%)	2,821 (18.5%)	67,555 (19.2%)
Quintile 4	47,126 (21.5%)	2,857 (21.3%)	22,404 (21.3%)	3,130 (20.6%)	75,517 (21.4%)
Quintile 5(most)	56,288 (25.6%)	3,327 (24.6%)	29,847 (28.3%)	4,225 (27.7%)	93,687 (26.6%)
Residential instability					
Quintile 1(least)	19,562 (9%)	1,271 (9.4%)	10,607 (10%)	1,633 (10.7%)	33,073 (9.4%)
Quintile 2	29,975 (13.7%)	1,814(13.4%)	14,666 (13.9%)	2,075 (13.6%)	48,530 (13.8%)
Quintile 3	39,798 (18.1%)	2,239 (16.6%)	18,228 (17.3%)	2,505 (16.4%)	62,770 (17.8%)
Quintile 4	48,684 (22.2%)	2,881 (21.4%)	23,170 (22%)	3,186 (20.9%)	77,921 (22.1%)
Quintile 5(most)	78,080 (35.8%)	5,122 (38.0%)	37,181 (35.6%)	5,647 (37.2%)	126,030 (35.7%)
Ethnic concentration					
Quintile 1(least)	50,313 (22.9%)	2,217 (16.5%)	21,266 (20.1%)	2,334 (15.3%)	76,130 (21.6%)
Quintile 2	46,945 (21.3%)	2,369 (17.6%)	20,611 (19.6%)	2,499 (16.5%)	72,424 (20.5%)
Quintile 3	43,276 (19.8%)	2,744 (20.3%)	20,192 (19.2%)	2,973 (19.5%)	69,185 (19.6%)
Quintile 4	39,510 (18.3%)	3,093 (22.9%)	20,050 (19.3%)	3,491 (22.9%)	66,144 (18.7%)
Quintile 5(most)	36,055 (16.7%)	2,904 (21.5%)	21,733 (20.6%)	3,749 (24.6%)	64,441 (18.4%)
Dependency					
Quintile 1(least)	21,397 (9.8%)	1,497 (11.1%)	12,131 (11.7%)	2,004 (13.2%)	37,029 (10.5%)
Quintile 2	29,083 (13.4%)	2,013 (14.9%)	15,035 (14.4%)	2,343 (15.4%)	48,474 (13.8%)
Quintile 3	35,226 (16.1%)	2,171 (16.1%)	17,770 (16.9%)	2,599 (17.1%)	57,766 (16.4%)
Quintile 4	44,616 (20.3%)	2,618 (19.4%)	21,487 (20.3%)	2,916 (19.1%)	71,637 (20.3%)
Quintile 5(most)	85,777 (39.2%)	5,028 (37.3%)	37,429 (35.5%)	5,184 (34.0%)	133,418 (37.8%)
HFRS					
Mean (SD)	5.30 (3.28)	17.5 (3.1)	6.5 (3.4)	17.9 (3.5)	6.7 (4.7)
Median	4.6	16.5	6.1	16.9	5.5
Range	(0-14)	(14-37)	(0-14)	(14-41)	(0-41)
IQR	(2.4-7.5)	(15.1-18.9)	(3.7-9.1)	(15.3-19)	(3-9.1)
ECI					
Mean (SD)	0.44 (0.49)	0.57 (0.49)	2.8 (1.02)	3.2 (1.3)	1.24 (1.3)
Median	0	1	2	3	1
Range	(0-1)	(0-1)	(2-12)	(2-11)	(0-12)
IQR	(0-1)	(0-1)	(2-3)	(2-4)	(0-2)
ALCW					
Mean (SD)	13.63 (27.8)	27.3 (49.3)	16.4 (30.1)	32.4 (58)	15.8 (31.7)
Median	7	13	8	15	7
Range	(1-1406)	(1-1400)	(1-805)	(1-1385)	(1-1406)
IQR	(3-14)	(6-29)	(4-17)	(7-34)	(3-16)

Table S3. Common ICD10 codes between ECI and HFRS

ICD-10 Category	HFRS[§]	ECI[†]	Common
A	A41, A04, A09	A52	-
B	B96, B95	B18, B20–B22, B24	-
C	-	C81–C85, C88, C96, C90, C77–C80, C00–C26, C30–C34, C37–C41, C43, C45–C58, C60–C76, C97	-
D	D64	D65–D68, D69, D50, D51–D53	-
E	E87, E86, E53, E16, E55, E05, E83	E10, E11, E12, E13, E14, E00–E03, E89, E66, E40–E46, E22, E86, E87, E52	E86, E87
F	F00, F05, F03, F01, F10, F32	F10, F11–F16, F18, F19, F20, F22–F25, F28, F29, F30, F31, F20, F31, F32, F33, F34, F41, F43	F10, F32
G	G81, G30, G20, G40, G31, G45	G04, G11, G80, G81, G82, G83, G10–G13, G20– G22, G25, G31, G32, G35–G37, G40, G41, G93, G62	G81, G20, G40, G31
H	H54, H91	-	-
I	I69, I67, I95, I63	I09, I11, I12, I13, I25, I42, I43, I50, I44, I45, I47–I49, I05–I08, I09, I34–I39, I26, I27, I28, I70, I71, I73, I77, I79, I10, I15, I85, I86, I98	-
J	J96, J18, J69, J22	J40–J47, J60–J67, J68, J70	-
K	K59, K26, K92, K52	K55, K70, K71, K72– K74, K76, K25, K26, K27, K28, K29, K70	K26
L	L03, L89, L97, L08	L94	-
M	M25, M19, M81, M79, M41, M80, M48, M15	M05, M06, M08, M12, M30, M31, M32–M35, M45, M46	-
N	N39, N17, N19, N18, N28, N20	N18, N19, N25	N18, N19
Q	-	Q23	-
R	R29, R31, R41, R26, R56, R40, R54, R55, R44, R94, R33, R69, R32, R45, R47, R02, R63, R13, R00, R79, R11, R50	R00, R47, R56, R63, R64	R47, R56, R63
S	S00, S06, S42, S80, S22, S72, S32, S09, S01, S51	-	-
T	T83	T82, T51	-
U	U80	-	-
W	W19, W18, W06, W10, W01	-	-
X	X59	-	-
Y	Y95, Y84	-	-
Z	Z50, Z75, Z60, Z22, Z87, Z74, Z93, Z99, Z73, Z91	Z45, Z95, Z49, Z99, Z94, Z50, Z71, Z72, Z72	Z50, Z99

[†]The first three characters of all ICD-10 codes used in ECI.
[§]The first three characters of all ICD-10 codes used in HFRS.
Note: ECI=Elixhauser Comorbidity Index, HFRS=hospital frailty risk score.

Table S4. Frequency of ECI Items Among Risk Groups

ECI Item	Group 1 (ECI<2, HFRS≤14)	Group 2 (ECI<2, HFRS>14)	Group 3 (ECI≥2, HFRS≤14)	Group 4 (ECI≥2, HFRS>14)
	219,057 (62%)	13,497 (3.8%)	105,321 (29.8%)	15,231 (4.3%)
Chronic heart failure	29,369	1,771	14,078	2,018
Cardiac arrhythmia	39,417	2,394	19,151	2,727
Valvular disease	6,115	377	2,885	396
Pulmonary circulation disorders	4,639	262	2,166	327
Peripheral vascular disorders	7,333	434	3,428	500
Uncomplicated hypertension	62,404	3,808	30,038	4,303
Paralysis	6,658	423	3,148	482
Other neurological disorders	15,094	923	7,358	1,074
Chronic pulmonary disease	23,559	1,389	11,145	1,705
Uncomplicated diabetes	13,196	842	6,382	880
Complicated diabetes	33,815	2,133	16,366	2,333
Hypothyroidism	6,622	412	3,195	469
Renal failure	15,784	958	7,505	1,087
Liver disease	3,120	219	1,562	218
Peptic ulcer disease excluding bleeding	1,091	59	491	62
AIDS/HIV	25	<6†	13	<6†
Lymphoma	2,054	131	939	144
Metastatic cancer	6,410	423	3,052	478
Solid tumor without metastasis	8,375	510	4,067	527
Rheumatoid arthritis / Collagen vascular disease	3,294	209	1,505	218
Coagulopathy	4,772	305	2,291	344
Obesity	2,796	163	1,325	187
Weight loss	10,927	704	5,326	783
Fluid and electrolyte disorders	35,011	2,179	16,772	2,415
Blood loss anemia	995	74	495	56
Deficiency anemia	5,056	271	2,474	385
Alcohol abuse	4,836	337	2,402	374
Drug abuse	766	51	387	56
Psychoses	2,153	133	1,033	137
Depression	10,633	657	5,044	737

†The real values cannot be revealed because of the agreement protocol

Table S5. Comparing models based on various forms of interaction between frailty and comorbidity

Models§	AIC
Model with Linear Interaction	
HFRS × ECI	477745.3
Model with Curvilinear Interaction	
B-Spline Basis (HFRS × ECI)	478450.6
Natural splines (HFRS × ECI)	478521.6
Orthogonal polynomial splines (HFRS × ECI)	478450.6

§The degree of the piecewise polynomial=3 for all models

Notes: ECI=Elixhauser Comorbidity Index, HFRS=hospital frailty risk score
Both ECI and HFRS are continuous variables

This table summarizes results regarding the fit of various logistic regression models containing various forms of interactions (including linear or curvilinear) between frailty and multimorbidity indices (using their continuous version), adjusted for the potential covariates. The models are compared in terms of AIC measure, which evaluates the goodness of fit. We used simple orthogonal polynomial splines, B-splines, and natural splines to implement the curvilinear interaction between frailty and comorbidity scores [12].

Table S6. Results of adjusted model stratified by short-moderate and long waiters to discharge

	Relative Risk Ratio (95% Confidence Interval)		
Outcome	**Direct Readmission**	**ED Readmission**	**Mortality**
Long Waiters[†]			
Group 2: MCC: No, Frailty: High	0.91 (0.69-1.13)	**1.09 (0.96-1.21)**	1.12 (0.83-1.41)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.19 (1.07-1.31)	1.40 (1.33-1.47)	1.50 (1.33-1.66)
Group 4: MCC: Yes, Frailty: High	1.39 (1.22-1.57)	1.60 (1.50-1.70)	1.67 (1.44-1.90)
Short-Moderate Waiters[§]			
Group 2: MCC: No, Frailty: High	1.01 (0.91-1.10)	**1.22 (1.16-1.27)**	1.04 (0.90-1.17)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.20 (1.16-1.24)	1.34 (1.31-1.36)	1.75 (1.70-1.80)
Group 4: MCC: Yes, Frailty: High	1.23 (1.14-1.32)	1.45 (1.40-1.50)	1.89 (1.78-1.99)
Original Study (short-moderate & long waiters)			
Group 2**:** MCC: No, Frailty: High	0.99 (0.91-1.06)	**1.19 (1.14-1.23)**	1.05 (0.95-1.15)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.20 (1.17-1.23)	1.34 (1.33-1.36)	1.73 (1.69-1.77)
Group 4: MCC: Yes, Frailty: High	1.26 (1.20-1.33)	1.48 (1.44-1.52)	1.83 (1.75-1.91)

[†]ALCW < Q3+1.5×IQR
[§]ALCW > Q3+1.5×IQR
Notes: ED= Emergency Department, MCC= Multiple Chronic Condition, ALCW: Alternate Level of Care Wait Time

APPENDIX C Finding optimal cutoff point for the HFRS dichotomization

The original study by Gilbert and colleagues categorized HFRS as low risk, L, (<5), moderate risk, M, (5–15), and high risk, H, (>15) [13]. Therefore, the cut-off points for the three categories of HFRS are 5 and 15. In this study, however, we opted to use a dichotomous version of HFRS, particularly for the ease of interpretations and interaction analysis. To impose minimal changes to the original categories (hence, the cut-off points) and maintain their order (i.e., the original L–M–H order), our dichotomization procedure boiled down into deciding whether the M category must be collapsed on the L category (hence, the resulting dichotomy of LM (low-to-moderate risk) vs. H (high risk)) or on the H category (hence, the resulting dichotomy of L (low risk) vs. MH (moderate-to-high risk)). This translates into choosing the cut-off point between 5 or 15. Our approach, therefore, optimizes two criteria simultaneously: interpretability and discriminatory power of the cut-off point. To this end, first, we examined how well the two cut-off options perform in terms of separating the resulting groups in terms of the outcomes using logistic regression. In doing so, we utilized the outcome-based stratification approach for finding the best optimal cut-point proposed by Williams [14]. We modified the approach for our categorical outcome. This algorithm is most appropriate when a threshold value already exists, as in our case. The modified approach consists of the following steps:

1) We defined a multinomial logistic regression with a dichotomous frailty score as the main variable for each candidate cut-off point.
2) We applied the bootstrapping procedure with 100 iterations to have approximately unbiased estimates of the effect sizes (i.e., relative risk ratio) and p-values.
3) We assigned a total score to each candidate cut-off point based on the corresponding p-value and the effect size for each bootstrap iteration. If the candidate cut-off point achieves

the largest effect size and lowest p-value, it is considered as the winner cut-off point. If none of the candidate cut-off points dominate others in terms of two criteria (i.e., the effect size and p-value), we assign zero and report no winner.

4) We ranked the candidates for each outcome (i.e., direct readmission, ED readmission, and death) based on their probability of domination (i.e., how many times the candidate beat its competitors).
5) We selected the candidate with the highest domination probability, averaged for the outcomes.

Our findings for the stage of the analysis reveal that the two groups of patients can be better separated by the cut-off point of 15 (Table S7).

Table S7. Comparing models by sensitivity on the cut-off points of HFRS (5 vs. 15)

Cut-off Point	Direct Readmission	ED Readmission	Death	Average Score
5	90%	0.0%	1.9%	31%
15	10%	100.0%	98.1%	**69%**
Winner				**15**

Note: In this analysis, the non-dominated scenarios were removed during iterations

Second, we conducted a sensitivity analysis to assess whether slight changes around the winner cut-off (i.e., 15) would perform better. We have developed multiple multinomial logistic regressions by changing the cut-off points from 13 to 17, i.e., two points greater and lower than 15, and repeated the algorithm. The results of the comparisons in terms of the total score are presented in Table S8, characterizing that the cut-off point of 14 yields better results.

Table S8. The results of sensitivity analysis on the cut-off points of HFRS (15 vs. 13, 14, 16,17)

Cutoff Point	Direct Readmission	ED Readmission	Death	Average Score
13	0.0%	0.0%	0.0%	0.0%
14	26.9%	100%	52.8%	**59.9%**
15	7.7%	0.0%	0.0%	2.6%
16	0.0%	0.0%	47.2%	15.7%
17	65.4%	0.0%	0.0%	21.8%
Winner				**14**

Note: In this analysis, the non-dominated scenarios were removed during iterations

APPENDIX D NUMERICAL DETAILS OF THE FIGURES

Table S9. Numerical details of Fig. 2

	Relative Risk Ratio (95% Confidence Interval)		
Outcome	**Direct Readmission**	**ED† Readmission**	**Mortality**
Group 2: MCC§: No, Frailty: High	0.99 (0.91-1.07)	1.19 (1.15-1.23)	1.05 (0.95-1.15)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.21 (1.18-1.23)	1.34 (1.32-1.36)	1.74 (1.70-1.77)
Group 4: MCC: Yes, Frailty: High	1.26 (1.19-1.33)	1.48 (1.44-1.52)	1.83 (1.75-1.91)
Sex (Female vs. Male)	0.83 (0.81-0.86)	0.89 (0.87-0.90)	0.72 (0.69-0.76)
Discharge destination (non-home vs. home)	1.72 (1.66-1.78)	0.71 (0.68-0.74)	1.15 (1.09-1.22)
Residency (Rural vs. Urban)	1.66 (1.62-1.70)	1.07 (1.04-1.09)	1.21 (1.15-1.27)
ALCW‡ (>14d vs. ≤14d)	0.90 (0.87-0.93)	1.05 (1.03-1.07)	0.95 (0.91-0.99)
Ethnic Concentration_Quartile2	0.97 (0.93-1.01)	1.01 (0.98-1.04)	1.02 (0.97-1.08)
Ethnic Concentration_Quartile3	0.89 (0.85-0.93)	1.05 (1.02-1.07)	0.99 (0.94-1.05)
Ethnic Concentration_Quartile4	0.72 (0.68-0.77)	1.06 (1.03-1.09)	1.02 (0.97-1.08)
Ethnic Concentration_Quartile5	0.65 (0.61-0.70)	1.11 (1.09-1.14)	1.09 (1.02-1.14)

†ED: Emergency Department
§MCC: Multiple Chronic Condition
‡ALCW: Alternate Level of Care Wait Time

Table S10. Numerical details of Fig. 3-a

	Relative Risk Ratio (95% Confidence Interval)		
Outcome	**Direct Readmission**	**ED† Readmission**	**Mortality**
Group 2: MCC§: No, Frailty: High	1.03 (0.92-1.14)	1.25 (1.18-1.32)	1.10 (0.95-1.25)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.22 (1.18-1.26)	1.28 (1.25-1.31)	1.58 (1.53-1.63)
Group 4: MCC: Yes, Frailty: High	1.31 (1.21-1.41)	1.44 (1.38-1.50)	1.69 (1.58-1.80)
Discharge destination (non-home vs. home)	1.68 (1.59-1.77)	0.72 (0.68-0.76)	1.08 (0.99-1.17)
Residency (Rural vs. Urban)	1.64 (1.59-1.69)	1.02 (0.98-1.06)	1.15 (1.06-1.24)
ALCW‡ (>14d vs. ≤14d)	0.89 (0.84-0.94)	1.03 (1.01-1.05)	0.90 (0.85-0.95)
Ethnic Concentration_Quartile2	0.95 (0.89-1.01)	1.03 (0.99-1.07)	1.03 (0.95-1.11)
Ethnic Concentration_Quartile3	0.89 (0.83-0.95)	1.07 (1.03-1.11)	0.98 (0.89-1.06)
Ethnic Concentration_Quartile4	0.73 (0.66-0.79)	1.09 (1.05-1.13)	0.99 (0.90-1.07)
Ethnic Concentration_Quartile5	0.66 (0.59-0.73)	1.14 (1.09-1.19)	1.10 (1.02-1.19)

†ED: Emergency Department
§MCC: Multiple Chronic Condition
‡ALCW: Alternate Level of Care Wait Time

Table S11. Numerical details of Fig. 3-b

	Relative Risk Ratio (95% Confidence Interval)		
Outcome	**Direct Readmission**	**ED† Readmission**	**Mortality**
Group 2: MCC§: No, Frailty: High	0.96 (0.86-1.06)	1.15 (1.09-1.21)	1.01 (0.87-1.15)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.18 (1.14-1.22)	1.40 (1.37-1.43)	1.90 (1.85-1.95)
Group 4: MCC: Yes, Frailty: High	1.23 (1.13-1.33)	1.51 (1.46-1.56)	1.97 (1.86-2.08)
Discharge destination (non-home vs. home)	1.76 (1.68-1.83)	0.70 (0.66-0.74)	1.24 (1.14-1.33)
Residency (Rural vs. Urban)	1.67 (1.62-1.73)	1.11 (1.07-1.15)	1.27 (1.19-1.35)
ALCW‡ (>14d vs. ≤14d)	0.91 (0.87-0.95)	1.05 (1.02-1.07)	1.00 (0.96-1.03)
Ethnic Concentration_Quartile2	0.99 (0.94-1.03)	1.00 (0.96-1.03)	1.02 (0.94-1.09)
Ethnic Concentration_Quartile3	0.88 (0.827-0.94)	1.02 (0.99-1.06)	1.01 (0.93-1.09)
Ethnic Concentration_Quartile4	0.72 (0.66-0.77)	1.03 (0.99-1.07)	1.06 (0.98-1.14)
Ethnic Concentration_Quartile5	0.65 (0.58-0.71)	1.09 (1.06-1.13)	1.07 (0.98-1.15)

†ED: Emergency Department
§MCC: Multiple Chronic Condition
‡ALCW: Alternate Level of Care Wait Time

Table S12. Numerical details of Fig. 4

	Relative Risk Ratio (95% Confidence Interval)		
Outcome	**Direct Readmission**	**ED† Readmission**	**Mortality**
7-days			
Group 2: MCC§: No, Frailty: High	0.75 (0.61-0.89)	1.20 (1.34-1.27)	1.16 (0.90-1.42)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.08 (1.03-1.13)	1.28 (1.26-1.30)	1.69 (1.59-1.79)
Group 4: MCC: Yes, Frailty: High	1.00 (0.88-1.12)	1.42 (1.37-1.48)	1.75 (1.55-1.96)
30-days			
Group 2: MCC§: No, Frailty: High	0.99 (0.91-1.06)	1.19 (1.14-1.23)	1.05 (0.95-1.15)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.20 (1.17-1.23)	1.34 (1.33-1.36)	1.73 (1.69-1.77)
Group 4: MCC: Yes, Frailty: High	1.26 (1.20-1.33)	1.48 (1.44-1.52)	1.83 (1.75-1.91)
90-days			
Group 2: MCC§: No, Frailty: High	0.63 (0.54-0.72)	1.18 (1.15-1.22)	1.15 (1.08-1.22)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.25 (1.22-1.28)	1.40 (1.38-1.41)	1.79 (1.77-1.82)
Group 4: MCC: Yes, Frailty: High	1.11 (1.04-1.18)	1.56 (1.52-1.58)	1.83 (1.78-1.90)
180-days			
Group 2: MCC§: No, Frailty: High	0.61 (0.53-0.69)	1.17 (1.14-1.20)	1.16 (1.10-1.22)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.27 (1.25-1.30)	1.41 (1.40-1.43)	1.81 (1.78-1.83)
Group 4: MCC: Yes, Frailty: High	1.11 (1.05-1.18)	1.57 (1.54-1.60)	1.89 (1.84-1.94)
270-days			
Group 2: MCC§: No, Frailty: High	0.60 (0.53-0.68)	1.17 (1.14-1.20)	1.18 (1.12-1.23)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.28 (1.26-1.31)	1.42 (1.41-1.43)	1.81 (1.79-1.83)
Group 4: MCC: Yes, Frailty: High	1.14 (1.08-1.20)	1.58 (1.55-1.62)	1.91 (1.86-1.96)
365-days			
Group 2: MCC§: No, Frailty: High	0.61 (0.54-0.68)	1.19 (1.16-1.23)	1.22 (1.17-1.28)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.30 (1.27-1.32)	1.44 (1.43-1.45)	1.82 (1.81-1.86)
Group 4: MCC: Yes, Frailty: High	1.18 (1.12-1.24)	1.62 (1.59-1.65)	1.94 (1.89-1.99)

†ED= Emergency Department
§MCC: Multiple Chronic Condition

Table S13. Numerical details of Fig. S2

	Relative Risk Ratio (95% Confidence Interval)		
Outcome	**Direct Readmission**	**ED† Readmission**	**Mortality**
Original Model:			
Non-home (Alternate care)	1.72 (1.66-1.78)	0.71 (0.68-0.73)	1.15 (1.08-1.22)
§**Sensitivity Model:**			
Home with support	0.66 (0.59-0.73)	1.13 (1.10-1.16)	1.35 (1.28-1.42)
Community or Long-term care	2.19 (2.13-2.25)	0.53 (0.51-0.56)	1.07 (1.01-1.14)

†ED: Emergency Department
§Definition of discharge destination changed from the dichotomous indicator to a three-level indicator (Reference is home with self-care for both models)

Table S14. Numerical details of Fig. S3

	Relative Risk Ratio (95% Confidence Interval)			
Outcome	**Direct Readmission (Original analysis)**	**ED† Readmission (Original analysis)**	**Readmission (combined)**	**ED Readmission (Excluding direct)**
Group 2: MCC§: No, Frailty: High	0.99 (0.91-1.06)	1.19 (1.14-1.23)	1.19 (1.15-1.23)	1.19 (1.14-1.23)
Group 3: MCC: Yes, Frailty: Low-to-Moderate	1.20 (1.17-1.23)	1.34 (1.32-1.36)	1.34 (1.33-1.36)	1.34 (1.32-1.36)
Group 4: MCC: Yes, Frailty: High	1.26 (1.20-1.33)	1.48 (1.44-1.52)	1.48 (1.45-1.51)	1.48 (1.44-1.52)

†ED: Emergency Department
§MCC: Multiple Chronic Condition

APPENDIX E R-CODE FOR CALCULATIONS OF HFRS

```
### Load Library
library(tidyverse)

### Upload Data
HFRS_ICD <- read.csv("HFRS_ICD.csv")
HFRS_Point <- read.csv("HFRS_Point.csv")
df=read.csv("DAD.csv")

### Filter Data
Comorbid_data <- df %>% select (id, days_to_admdate, dx10code1: dx10code25) %>%
mutate (visitID = paste (id, days_to_admdate, sep="_"))

### Calculate the HFRS

# n: number of HFRS ICD

for (i in 1: n) {
       id_ls <- Comorbid_data %>% filter_at (vars (dx10code1: dx10code25),
       any_vars(substr (.,1,3) == HFRS_ICD [i])) %>%
       select (id, visitID, dx10code1: dx10code25) %>%. $visitID
       var_name <- as. character (HFRS_ICD [i])
       Comorbid_data<-Comorbid_data %>% mutate (!!var_name: =ifelse (visitID %in%
       id_ls,1,0))
}
HFRS_score= as.matrix(Comorbid_data %>% select
                           (HFRS_ICD [1]: HFRS_ICD [n]))%*% as.matrix (HFRS_Point)
```

REFERENCES

[1] Kahlon S, Pederson J, Majumdar SR, Belga S, Lau D, Fradette M, et al. Association between frailty and 30-day outcomes after discharge from hospital. Cmaj 2015;187:799–804.
[2] Qiukui H, Zhou L, Biao D, Yang M, Birong D, Yuquan W. The role of frailty in predicting mortality and readmission in older adults in acute care wards: a prospective study. Sci Rep Nat Publ Group 2019;9.
[3] Li Y, Pederson JL, Churchill TA, Wagg AS, Holroyd-Leduc JM, Alagiakrishnan K, et al. Impact of frailty on outcomes after discharge in older surgical patients: a prospective cohort study. CMAJ 2018;190:E184–90.
[4] Vidán MT, Blaya-Novakova V, Sánchez E, Ortiz J, Serra-Rexach JA, Bueno H. Prevalence and prognostic impact of frailty and its components in non-dependent elderly patients with heart failure. Eur J Heart Fail 2016;18:869–75.
[5] Tran DT, Tu JV, Dupuis J-Y, Bader Eddeen A, Sun LY. Association of frailty and long-term survival in patients undergoing coronary artery bypass grafting. J Am Heart Assoc 2018;7:e009882.
[6] Hatcher VH, Galet C, Lilienthal M, Skeete DA, Romanowski KS. Association of clinical frailty scores with hospital readmission for falls after index admission for trauma-related injury. JAMA Netw Open 2019;2:e1912409–e1912409.
[7] Hewitt J, Carter B, Vilches-Moraga A, Quinn TJ, Braude P, Verduri A, et al. The effect of frailty on survival in patients with COVID-19 (COPE): a multicentre, European, observational cohort study. Lancet Public Health 2020;5:e444–51. https://doi.org/10.1016/S2468-2667(20)30146-8.
[8] Nguyen QD, Wu C, Odden MC, Kim DH. Multimorbidity patterns, frailty, and survival in community-dwelling older adults. J Gerontol Ser A 2019;74:1265–70.
[9] Ritt M, Ritt JI, Sieber CC, Gassmann K-G. Comparing the predictive accuracy of frailty, comorbidity, and disability for mortality: a 1-year follow-up in patients hospitalized in geriatric wards. Clin Interv Aging 2017;12:293.
[10] Hanlon P, Nicholl BI, Jani BD, Lee D, McQueenie R, Mair FS. Frailty and pre-frailty in middle-aged and older adults and its association with multimorbidity and mortality: a prospective analysis of 493 737 UK Biobank participants. Lancet Public Health 2018;3:e323–32.
[11] Serina P, Lo AX, Kocherginsky M, Gray E, Lindquist LA, Post LA, et al. The Clinical Frailty Scale and Health Services Use for Older Adults in the Emergency Department. J Am Geriatr Soc 2020.
[12] Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. BMC Med Res Methodol 2019;19:1–16.
[13] Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. The Lancet 2018;391:1775–82.
[14] Williams BA. Finding optimal cutpoints for continuous covariates with binary and time-to-event outcomes 2006.

Chapter 4

A Data-Driven Approach to Address Care Fragmentation Among Older Adults: Prediction, Decision-Making, and Fairness Considerations

Ghazalbash, S., Zargoush, M., Verter, V., Guilcher, S., Kuluski, K. (2022). "Data-Driven Decision-Making Modeling of Continuity of Care: Predictive and Economic Investigations of Care Fragmentation in the Aging Population." Prepared to submit to "Production and Operation Management" journal.

Preface

The third article in this dissertation, entitled "A Data-Driven Approach to Address Care Fragmentation Among Older Adults: Prediction, Decision-Making, and Fairness Considerations," is targeted for submission to the Production and Operations Management (POM) journal. In this research, S. Ghazalbash and M. Zargoush have made substantial contributions to the concept, research design, and problem formulation. Overseen by M. Zargoush, data analytics were conducted by S. Ghazalbash, who also drafted the initial version of the manuscript. All authors contributed to interpreting the results and revised the manuscript for intellectual content. V. Verter provided critical feedback and helped shape the research. S. Guilcher and K. Kuluski provided insights regarding the clinical and policy implications of the results. M. Zargoush oversaw the entire study.

Abstract

A disruption in the continuity of care (COC), which refers to Care fragmentation (CF), is a major source of healthcare systems' inefficiency and occurs when care is spread over different providers. This challenge becomes more problematic among older patients as they are often medically complex with multifaceted needs and require transfer across multiple care settings, hence, the need for efficient care coordination. One of the key aspects of CF is fragmented readmission—that is, readmission to a hospital different from the initial hospital. One strategy for reducing fragmented readmission is predicting CF and imposing targeted interventions on patients with a high risk of CF. Despite the availability of large data and exponential advances in computational power, existing studies have not thoroughly investigated the predictive–prescriptive approach to provide actionable insights regarding fragmented readmission. This is mainly due to challenges with using machine learning (ML) predictions for clinical decision-making, including i) the complexity of explicitly incorporating predictive information for decision-making and b) concerns with the fairness implications of such a decision-making paradigm. This study proposes a data-driven predictive–prescriptive analytics framework that leverages ML using massive, longitudinal data to derive optimal CF intervention policies while addressing disparities in decision-making. We construct predictive ML-based models and estimate personalized costs utilizing one million observations collected over 13 years in Ontario, Canada's largest province. To examine the real-world implications of the ML-based strategy on cost and fairness, we apply the proposed framework to seven preventive interventions currently used in clinical practice. The framework in this study outperforms non-targeted implementation of the existing clinical strategies and provides actionable insights into managing CF. It also mitigates discriminatory decisions and offers significant financial savings compared to existing clinical strategies. In Ontario, implementing the

existing programs based on the proposed ML-based decision-making framework (as opposed to the random strategy) leads to less cost, contributing to potential $3.6 million to $5.2 million savings per year for 100,000 patients. Our proposed decision-making paradigm supports decision-making and resource planning toward a targeted allocation at the systems level and informs actions that affect patient-centered care transitions at the service level. It can also facilitate shared decision-making among aging people, their families, and care providers.

Keywords: *Care fragmentation; machine learning; fairness; older adults; non-index readmission; continuity of care, predictive–prescriptive analytics.*

1. Introduction and Motivation

A disruption in the continuity of care (COC), which refers to care fragmentation (CF), is one of the main causes of the healthcare system's inefficiency as it compromises care coordination among providers. It leads to increased healthcare costs (Rosenberg and Zulman 2020) and reduced quality of care (Hirji et al. 2020). CF also contributes to patient dissatisfaction and the progression of patient comorbidities (Hirji et al. 2020; Juo et al. 2019a). CF becomes more serious among older patients, who are often medically complex with multiple chronic conditions and geriatric syndromes. Therefore, they need to visit various care providers and transfer across multiple care settings, requiring efficient core coordination (Brooke 2020). Consequently, maintaining COC is a priority for clinicians and policymakers and an indicator of the quality of care, particularly for older patients (Barker et al. 2017; Hirji et al. 2020).

One of the critical aspects of COC is uninterrupted patient visits with the same provider (Saultz 2003), such as readmission to the same hospital (Haggerty et al. 2003). This COC aspect is important because hospitals are more likely to have shared data (physical and electronic) and administration systems to facilitate sharing patient critical information, such as admission, medical history, complications, progress notes, diagnostic test results, and in-hospital interventions. Therefore, maintaining interaction with the same hospital during the care journey is associated with greater formal and informal informational continuity across providers and improved quality of care (McAlister et al. 2017; Senot 2019). Information sharing among care providers facilitates collaborative care implementation and shared decision-making by interdisciplinary teams, resulting in safe clinical practice and better health outcomes, particularly among older patients with medically complex needs (Burm et al. 2019). On this note, fragmented hospital readmission refers to readmission to a non-index hospital, i.e., a hospital different from the initial one.

Fragmented readmission is a prevalent global issue affecting many healthcare systems, particularly in North America (Kaltenborn et al. 2021; McAlister et al. 2017). For example, in the United States, nearly one-third of Medicare beneficiaries are readmitted to a different care setting in a 1-year period (Kaltenborn et al. 2021). Recent studies have demonstrated the negative impact of fragmented readmission on post-discharge outcomes, including mortality and readmission, for a variety of patient populations, such as surgical (Juo et al. 2019a), orthotopic liver transplantation (Kothari et al. 2017), traumatic injury (Passman et al. 2020), emergency general surgery (McCrum et al. 2020), cardiac surgery (Hirji et al. 2020), and complex cancer surgery populations (Brauer et al. 2021). Fragmented readmission also imposes substantial financial burdens. The reported average cost difference between fragmented and nonfragmented readmissions ranges from $270 to $22,000 per patient (Snow et al. 2020).

One strategy for reducing fragmented readmission is to predict the occurrence of this event among patients and impose targeted interventions on those at a higher risk of fragmented readmission (Agha et al. 2019; Hirji et al. 2020). Given the exponential advances in data collection, data storage, and computational power over the past decade, using machine learning (ML) to solve problems in an informed, evidence-based fashion has gained great appeal (Chekroud et al. 2021). In health care, ML applications are also becoming popular (Doupe et al. 2019) because of their premise of predicting health outcomes with greater accuracy than traditional statistical methods (Chekroud et al. 2021). However, the existing literature has not thoroughly investigated ML approaches to provide actionable insights into the issue of fragmented readmission due to the challenges with using ML for targeted interventions and clinical decision-making, as discussed below.

First, the main premise of most ML-based studies is mainly predictive analytics—that is, ensuring high predictive performance. However, these approaches are less clear about how the predictions can be explicitly utilized for prescriptive analytics—that is, optimal decision-making—or their real benefits and practical insights in terms of financial cost savings, resource allocations, and patient quality of life (Morse et al. 2020; Teo et al. 2021). In the context of CF, the challenge is to identify patients for targeted preventive interventions (to minimize their risk of CF) and do this optimally in a resource (e.g., budget)-constrained environment while maintaining patient quality of care. In the current clinical practice and devoid of such predictive information, the common approaches include ad hoc strategies, such as "*random strategy*" or its extreme cases ("*treating all*," "*treating none*"), which are prone to sub-optimality. Considering the numerous promising interventions for preventing fragmented readmission, one key decision is the optimal choice of targeted interventions, given their implementation cost and effectiveness. Utilizing ML predictive information to identify patients at risk for fragmented readmission and explicitly considering the cost implications of ML-based targeted interventions can be an innovative, evidence-based, and practical solution to the above challenge.

Another challenge with ML-based decision-making is controlling bias toward protected groups (e.g., women, people of color, or patients of low socioeconomic status) and the resulting inequities. Such bias can lead to potential delays in providing the necessary care or treatment to these sensitive groups, resulting in lower quality of care (Ghassemi and Nsoesie 2022; Seyyed-Kalantari et al. 2021). To protect sensitive groups of people, anti-discrimination laws have been established in many countries, such as the United States (Fu et al. 2021), Croatia, Germany, Poland, Slovenia (Bielińska et al. 2022), and Canada (Hassen et al. 2021). These laws are particularly related to the provision of healthcare services. In addition, Healthy People 2020 has emphasized "achieving

health equity" and "eliminating disparities" in its mission (Rajkomar et al. 2018). The potential of automated ML-based systems to make equitable decisions in healthcare has inspired significant research efforts (MIT News 2021; Seyyed-Kalantari et al. 2021). However, the assumption that decisions made by automated models are always fair may not be correct. Although using ML-based models could revolutionize health care and medicine by improving the process, care, and ultimately health outcomes, they could also replicate the existing social bias and consequential harm. This fear has slowed down ML applications for clinical decision-making (Richardson 2022). Recent studies have highlighted the potential algorithmic bias that may produce discriminatory outcomes, exacerbating inequities among protected groups (Fu et al. 2021; Ghassemi and Nsoesie 2022). Specifically, healthcare data are subject to bias stemming from clinical data collection or measurement devices and inherent human bias in diagnosis or treatment (Ghassemi and Nsoesie 2022). Data are not uniformly distributed between protected and regular groups. Certain socioeconomic groups, such as black patients, minority groups, or female patients, are usually underrepresented in healthcare data because of biological or nonbiological variation (Kallus et al. 2022; Rajkomar et al. 2018; Wiens et al. 2019). ML algorithms trained on such biased data can perpetuate (or even magnify) the healthcare disparities already present in the data or produce new ones if left unchecked, potentially leading to discriminatory decisions about patient's access to care and raising ethical concerns (Ferrer et al. 2021; Rajkomar et al. 2018). A successful design and implementation of ML-based decision-making require addressing fairness concerns (Richardson 2022). By targeting fairness as a central consideration in the model design, deployment, and evaluation, one could potentially steer ML-based decision-making toward fair behavior and ensure equities (Rajkomar et al. 2018).

To address these gaps, this study aims to accomplish the following:

(i) Developing a competitive ML-based prediction model to identify patients at risk of fragmented readmission.

(ii) Developing a decision-analytics framework and intuitive optimal policies to illustrate how ML predictions can be used for targeted interventions in real clinical practices via extensive comparison with random intervention strategies.

(iii) Examining the fairness implications of the developed ML-based decision-making framework to ensure parity among protected groups.

Contributions. Our study contributes to the broader COC literature and ML-based healthcare decision-making in the following ways. First, we utilize a rich set of longitudinal data collected over a decade with approximately one million unique observations to develop ML-based predictive models for CF. Aside from its rich breadth and depth, a unique feature of our data is that they include hospitalization costs at the patient visit level, allowing us to estimate personalized data-driven hospital expenses and the resulting personalized optimal policies. To the best of our knowledge, this is among a few studies in the Operations Research/Management Science (OR/MS) domain to utilize data-driven cost estimations for clinical decision-making. The data were collected from more than 18 acute care facilities, and hence, they can address the available challenges of limited generalizability. Second, we incorporate these data to develop competitive prediction models that address two challenges of the existing ML-based predictive studies, as discussed above, i.e., a) the need for fair decisions about underserved patients negatively influenced by algorithmic bias (Fu et al. 2020) and b) a lack of cost-based decision analytics framework to evaluate the clinical impact of utilizing ML predictions on healthcare decision-making (Teo et al. 2021). Although each of these challenges has been investigated *separately* in the current literature, our study aims to overcome both challenges simultaneously to provide more

robust, practical, and holistic insights. Our data-driven optimization framework yields the optimal range of intervention costs in which an ML-based strategy can provide financial benefits compared to the current strategies. Third, from a more practical standpoint, we provide a clinical decision-making framework that assists both service- and system-level decision-makers. At the service level, the framework suggests an optimal screening policy for targeting patients at a high risk of CF while balancing the tradeoff between the cost of wrong screening (due to ML false predictions) and potential cost savings related to the current random strategies. At the system level, it assists policymakers in selecting optimal preventive interventions, given the available budget as well as intervention costs and effectiveness for varying degrees of risk aversion. Moreover, our recommendations can be tailored to patient attributes, providing a personalized recommendation framework. Fourth, to our knowledge, we are the first to provide a decision-making framework that addresses fairness in both the predictive and prescriptive senses. To this end, we combine the notion of algorithmic bias with a need-based resource allocation philosophy to characterize fair decisions. Finally, we contribute to aging research in light of its paucity in the OR/MS domain despite the growing population of older adults globally and the enormous cost implications of this population shift, which further necessitates optimal management of care, informed decision-making, and evidence-based policymaking for older adults. In Canada, around 50% of hospital expenditures are attributed to older adults, although this group represents only 14% of the Canadian population (Fox et al. 2013). Overall, this research lies at the intersection of analytics, healthcare management, and aging research.

2. Related Works

This study contributes to the COC and aging literature on the context side and ML-based predictive-prescriptive analytics and fair decision-making on the methodology side. Our review of related works focuses on the healthcare OR/MS literature after 2015.

Continuity of Care. COC is a multidimensional concept with two core notions: 'continuous caring relationship' and 'seamless service' (Gulliford et al. 2006; Lotz 2019). According to the Canadian Health Services Research Foundation, three types of COC can be defined—namely, informational continuity, relational continuity, and managerial continuity, which depend on the type and setting of care (Haggerty et al. 2003). Most studies focused on relational continuity—which refers to the ongoing relationship between patients and care providers— and have investigated its impact on patient outcomes and service utilization (Kajaria-Montag et al. 2021). Ahuja et al. (2019) showed that glycemic variability partially mediates the relationship between relational COC and system utilization metrics, whereas COC lowers glycemic variability through increased medication adherence. A follow-up study examined how diabetic patients' care continuity with their primary care physicians improves their health outcomes, such as length of stay (LOS), and 30-day readmission (Ahuja et al. 2020). They found a curvilinear relationship between relational COC and health outcomes, but on average, increasing COC was associated with improved outcomes across patients. Queenan et al. (2019) found that using technology-enabled relational COC, including telemonitoring and periodic telephone calls, and its interaction with patient engagement in care decisions reduce hospital readmissions. They also investigated how ML predictions of patient engagement can inform care providers about patients' needs and improve their COC. In a seminal study, Senot (2019) differentiated between relational and informational COC and defined a COC model that seeks to integrate care episodes at the individual-provider level (i.e., continuity

of referring provider), the accountable care organization (i.e., networks of providers) and across providers (i.e., continuity of physical location). The latter refers to the extent to which the episodes of care are situated in a single physical location, such as the same hospital or the same clinic, to improve care integration via common physical, electronic, and managerial systems. The study investigated how the integration of care episodes relates to the risk of readmission among patients with heart failure. It found that three care continuity mechanisms—that is, the individual referring provider, continuity of physical location, and continuity of accountable care organization—are associated with a significantly lower risk of patient readmission.

Our study focuses on the continuity of physical locations, which can promote the information. Unlike the above studies that measure the COC using the existing indices to investigate its association with outcomes, our study aims to predict the fragmentation of continued care in terms of physical location—namely, readmission to any non-index hospital). We capitalize on these predictions to design optimal preventive strategies, save healthcare costs, and assess their fairness.

Aging Research. Existing OR/MS studies have advanced aging research from different perspectives. Kong et al. (2022) investigated how the degree of standardization across service chains in different operational dimensions (i.e., customer mix, service offering, and service delivery) affects nursing home performance outcomes (e.g., financial performance, clinical outcome, and resident welfare). Mohliver and Ody-Brasier (2022) examined the association between organizational religious affiliation and violations of care standards in nursing homes. Jin et al. (2022) studied how a Medicare reimbursement rule affects elderly patients' discharge to a skilled nursing facility (SNF), leading to the overuse of SNFs and subsequent hospital readmission. Lu et al. (2021) proposed a theoretical model to assess the association between public-private payers' competition and service quality (measured as the 30-day rate of falls) in US nursing homes.

Focusing on elderly patients, Zychlinski et al. (2020) have formulated joint modeling of the hospital-geriatric institution to address bed allocation in geriatric wards by optimizing the number of geriatric beds while minimizing the costs associated with bed operations. Finally, Lu and Lu (2017) investigated how mandatory overtime laws for nurses affect the quality of care services at nursing homes. In contrast to these studies, which focused either on the predictive or prescriptive aspect of analytics, we developed a predictive–prescriptive framework that aims to improve older adults' care. Moreover, to our knowledge, none of the previous studies addressed the problem of CF among the older population, nor did they investigate the value of optimal ML-based strategies for preventive interventions compared to existing clinical practices.

ML-Based Decision Making in Healthcare. Studies using ML methodologies in healthcare OR/MS can be categorized into problems at the patient, clinic/hospital, and policy levels (Mišić and Perakis 2020). Regarding patient-level problems, Bertsimas et al. (2016) used predictive information to formulate an optimization model to identify the optimal chemotherapy regimens for gastric cancer. Bastani and Bayati (2020) developed a data-driven decision-making algorithm based on LASSO (least absolute shrinkage and selection operator) estimator to learn an optimal dosing strategy for warfarin prescriptions based on patients' clinical and genetic features. The developed predictive–prescriptive model outperforms the benchmark policy used in practice. Furthermore, Merdan et al. (2017) proposed a data analytics method to facilitate decision-making by urologists to optimally detect metastatic prostate cancer. Their model significantly reduced both false and missed metastatic imaging. Bjarnadottir et al. (2018) proposed a web-based decision support system that assists physicians with decisions about colorectal cancer treatments based on the ML-predicted patients' survival.

Regarding the clinic/hospital-level problems, Ang et al. (2016) proposed a model to predict wait time until starting treatment for low-acuity patients during emergency department (ED) triage. The model helps decision-makers choose the patients who will benefit from expedited treatment. Bertsimas et al. (2021) developed ML-based models to predict several aspects of patient flow, including short-term discharge volumes, long-stay patients, discharge disposition, and intensive care needs, to inform daily bed placement decisions. Furthermore, Samorani et al. (2021) proposed an appointment scheduling model that optimally assigns appointment requests based on the predicted probabilities of the patients' risk-of-show while minimizing schedule cost and racial disparity. Regarding policy-level problems, Kamalzadeh et al. (2021) proposed a screening strategy that involves combining predictive analytics for predicting risk of diabetes, hidden Markov to generate transition and emission rates, and the Markov decision process to derive optimal screening decisions. Their developed optimal policies outperformed the existing ad-hoc approach used in clinical practice.

We contribute to this stream of literature by developing a data-driven predictive–prescriptive analytics framework that leverages big data and ML to derive optimal intervention policies for preventing CF while addressing disparities in the decision-making process. To this end, we use ML with electronic health records of around one million inpatient visit observations to predict the individualized risk of CF. We then use this information to design optimal preventive strategies for patients at high risk of fragmented readmission. Our decision-making framework optimally trades off the cost of false/missed screening and the potential savings compared with the common clinical practice. It then identifies the optimal range of intervention costs through which the ML strategy is optimal.

Fair Decision-making. Algorithmic fairness has received considerable attention in different contexts, such as fair pricing/price fairness (Chen et al. 2022; Cohen et al. 2022), mortgage lending (Bartlett et al. 2022), recidivism prediction instruments (Chouldechova 2017), the criminal justice system (Corbett-Davies et al. 2017), decision making in firms and institutions (Fu et al. 2021), and healthcare (Kallus et al. 2022). These studies provide frameworks for assessing discrimination in decision-making under various fairness constraints. Earlier studies have also highlighted concerns about potential algorithmic bias raised by ML-based decision-making tools. In response, a growing body of literature has developed "fair" ML algorithms that attempt to mitigate discriminatory outcomes against protected groups. Shimao et al. 2021, 2022 defined a new notion of fairness in algorithmic bias— called "strategic best-response fair"— that address both disparity in the prediction results and disparity in the behavior of prediction subjects. Fu et al. 2021 also featured the importance of considering the strategic behavior of subjects in the context of fair ML. They provided a framework to compare the effect of two well-known interpretations of discrimination—that is, treatment disparity and impact disparity— on the firm's profit and the defaulters when considering the firm's strategic role in algorithmic decision-making. Kallus et al. 2022 studied the fairness of algorithmic decisions—defined as demographic parity and classification parity— from the perspective of data combination when some protected attributes are unobserved in the data. They applied their approach in the real case studies of lending and healthcare. Seyyed-Kalantari et al. 2021 examined false positive rate parity in chest X-ray classification models produced using computer vision techniques. They investigated the algorithmic bias for both subgroups and intersectional groups—that is, patients' membership to more than one under-deserved group —and concluded that algorithmic bias among protected groups interacts with each other. Qi et al. 2021 developed a new approach— inspired by ML

fairness metrics in AI applications—to measure disparities in randomized clinical trials (RCTs). They assessed the representativeness of subjects in RCTs and identified under-represented subgroups that contribute to different enrollment rates in RCTs, hence health inequity.

Most studies in this stream have investigated technical and practical aspects of algorithmic bias in a predictive sense. However, we provided a framework that addresses the issue of disparities among protected groups from both predictive and prescriptive views. Although we resembled the common fairness definitions in the fair ML literature, we combined the notions with a need-based resource allocation philosophy to characterize fair decisions in the context of CF.

3. Materials and Methods

3.1. Proposed Methodology

Figure 1 summarizes the whole predictive-prescriptive framework. After data preparation (i.e., transformation, missing data imputation, and resampling) in Step 1, we empirically investigated the negative impacts of fragmented readmission. To this end, we used multinomial logistic regression to assess the association between fragmented readmission and three outcomes, i.e., resource utilization, LOS, and death following readmission (Step 2). The details of this analysis are summarized in Appendix A. To initiate the predictive analytics, hyperparameters were tuned for each ML algorithm to identify the best prediction algorithm (Step 3). A stacked generalization approach was also used to develop an ensemble model based on multiple ML algorithms. Further, we evaluated the performance of the best models on the test set (details in Section 4.2). Next, we obtained the Pareto optimal set and plotted the Pareto front line to compare the constructed predictive models on multiple performance measures and identify the nondominated predictive model choice (details in Appendix C3). This analysis shows the best achievable tradeoff between various measures of model performance. We developed the ML-based decision analysis in Step 4,

where we conducted a comprehensive cost evaluation to determine the economic benefits of ML-based decision-making (details in Section 4.3). Finally, in Step 5, we evaluated algorithmic fairness to ensure nondiscriminatory decision-making (details in Section 4.4).

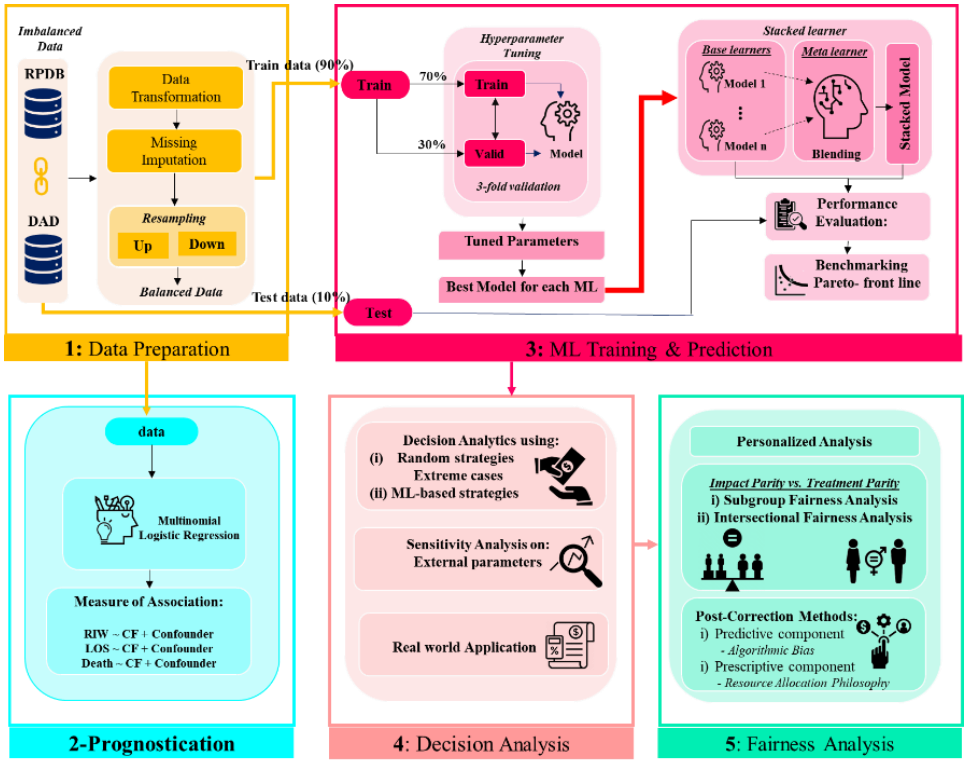


Figure 1. The methodological framework used in this study

3.2. Data Sources

We carried out a retrospective cohort study of individuals aged 65 and older in Ontario, Canada's largest province, between April 2004 and March 2016. Two databases housed in the Institute for Clinical Evaluative Sciences (IC/ES) were used to create the study data. The Discharge Abstract Database (DAD) provides information on patients' health status and use of health services. The patients' sociodemographic characteristics were extracted using the Registered Persons Database (RPDB). The initial dataset contained 6,039,054 observations from 1,741,830 patients. For the predictive analytics, we excluded patients with the following characteristics: (i) patients who died during the index hospital visits, as there was no risk of readmission for them; (ii) patients who

were transferred to another acute care hospital; (iii) patients who left the hospital against medical advice; (iv) patients who had planned readmission; and (v) patients who did not return from a pass, as they did not fit our definition of unplanned readmission. A total of 3,748,775 observations were found eligible to be included in this study. We further restricted our cohort to include only patients with at least one readmission within 30 days of discharge, yielding 953,818 unique observations for model development. Appendix D displays the data eligibility flow diagram in our study. This study was approved by the ethical board of the Hamilton Integrated Research Ethics Board (HiREB) in Ontario.

3.3. Data Descriptions and Preparation

Candidate predictor selection was guided by prior work and data availability. Predictor variables included the following: (i) patient demographics (age, gender), (ii) socioeconomic status (marginalization and rural/urban residency), (iii) clinical variables (multimorbidity burden, frailty burden, the most common primary diagnosis for hospitalization, main patient services, and the most common interventions), and (iv) administrative variables (hospital LOS, mode of readmission, discharge disposition [routine vs. other], admission to special care unit [SCU], SCU LOS, hospitalization history indicator, and patients with delayed discharge indicator).

Marginalization was measured using the Ontario Marginalization Index (Matheson and Van Ingen 2016), which considers material deprivation, residential instability, ethnic concentration, and dependency. These items are measured on a quantile scale from Q1 to Q5, where Q5 represents the most severe level of marginalization. The multimorbidity burden was measured using two common comorbidity indices—namely, the Charlson–Deyo Comorbidity Index (CDCI; Deyo et al. 1992) and the Elixhauser Comorbidity Index (ECI; Elixhauser et al. 1998). In addition, the frailty burden was measured using the Hospital Frailty Risk Score (HFRS; Gilbert et al. 2018).

Two alternative comorbidity indices were also assessed based on Quan's adaptations of the CDCI (Quan et al. 2011) and Van Walraven's adaptation of the ECI (van Walraven et al. 2010). The most common primary diagnosis for hospitalization was assessed using the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) codes. Acute myocardial infarction, heart failure, pneumonia, chronic obstructive pulmonary disease (COPD), osteoarthritis of the knee, and urinary system disorders were the top primary diagnoses, referring to patients' most resource-intensive conditions. Since recent studies have demonstrated that sepsis is one of the top causes of readmission, we also identified patients with sepsis (Lawson et al. 2021). The most frequent patient services were general medicine, cardiology, respirology, general surgery, orthopedic surgery, and urology. The 14 flagged interventions—the high-cost interventions proposed by the Canadian Institute for Health Information (CIHI)—were the use of a feeding tube; parenteral nutrition; chemotherapy; radiotherapy; use of a vascular access device; dialysis; cardioversion; cell saver; paracentesis; pleurocentesis; tracheostomy; invasive ventilation (greater than 96 hours vs. less than 96 hours); heart resuscitation. We also considered the most common interventions assessed by the comprehensive ambulatory classification system, including percutaneous transluminal coronary vessel intervention, joint replacement, prostate resection and biopsy, lower urinary tract intervention, and interventions not generally ambulatory.

The hospital LOS was separated by alternate level of care (ALC) wait time and acute LOS. The ALC wait time captures the number of days a discharge was delayed after a patient is medically ready for discharge. The mode of readmission is defined as readmission via ED or direct readmission to the hospital (e.g., from a clinic, doctor's office, or day surgery). The medical intensive care nursing unit, surgical intensive care nursing unit, combined medical/surgical intensive care nursing unit, coronary intensive care nursing unit medical, step-down medical unit,

and step-down surgical unit were the most common types of special care units that patients are admitted to during hospitalization. Finally, the hospitalization history was defined based on the occurrence of at least one acute hospitalization 90 days prior to admission. The outcome was non-index unplanned readmission (i.e., fragmented readmission), defined as readmission within 30 days to a different hospital from the one at which the previous admission occurred (Hirji et al. 2020).

The possibility of multicollinearity was examined through the predictors' variance inflation factor. Missing values for categorical variables were imputed from the mode—that is, the most probable known value (Maimon and Rokach 2005). Regarding data transformation, we constructed comorbidity indices using the "icd" package in R (version 3.6). We discretized the age variable into two intervals using the binning method, with intervals of ≤79 and ≥80. The discharge destination was dichotomized based on whether the patient was discharged home with self-care or to other destinations (e.g., home with support, long-term care, rehabilitation).

3.4. Predictive Analytics: Building and Evaluating ML algorithms

For an extensive evaluation of the ML predictive performance, we developed seven predictive models using the classification and regression tree (CART), random forest (RF), extreme gradient boosting (XGB), logistic regression (LR), naïve Bayes (NB), adaptive boosting (ADA), and a stacked generalization learner (Stacked Ensemble) approaches under the same data structure/preparation, model training, and assessment procedures. The stacked generalization is an ensemble model that blends multiple submodels (RF, XGB, and ADA in our study) to predict the outcome. Then, a meta-learner (LR in our study) took the outputs from sub-models to learn how much weight each model's predictions should receive to make a better output prediction (Chatzimparmpas et al. 2020). We specifically used tree-based ML algorithms as they have

performed well in clinical and administrative practice, such as early identification of hospitalization (De Hond et al. 2021), prediction of patients at risk of stroke at emergency department triage (Sung et al. 2021), and predicting triage level (Jiang et al. 2021). They do not require statistical assumptions such as the absence of multicollinearity, they are simple and robust to outliers, and can capture complex non-linear relationships (De Hond et al. 2021; Fernandes et al. 2021; Song et al. 2021; Suresh et al. 2020). Some studies have also shown that LR can perform as well as ML algorithms in predicting acute kidney injury (Song et al. 2021) or patient mortality (Cowling et al. 2021). Therefore, we considered this algorithm, too. All models were trained and evaluated using R 3.6.0 and package "*mlr3"*.

For predictive performance assessment, the dataset was randomly divided into a training set (60%), a validation (tuning) set (30%), and a test set (10%) to train and evaluate the ML algorithms. All models were evaluated using a fivefold cross-validated area under the curve (AUC) to determine the optimal hyperparameter values and avoid overfitting. To optimally tune the hyperparameters of the ML algorithms, we used the grid search algorithm. We also reported the true-positive rate (sensitivity/recall), true negative rate (specificity), positive predictive value (PPV; precision), F1 measure, and accuracy. We handled the challenge with imbalanced distributions of outcome classes by data resampling techniques, including down- and up-sampling, during the training. To balance class distribution, undersampling removes instances with the majority class, whereas oversampling replicates some instances with the minority class (Fotouhi et al. 2019; Leevy et al. 2018). We also assessed the calibration of the predictive models. The calibration curve evaluates how well the predicted probabilities agree with the observed probabilities (Van Calster et al. 2018).

For each algorithm, we defined a combination profile for hyperparameter values. Then, we evaluated the performance of ML algorithms under each profile using the fivefold cross-validated AUC and selected the one yielding the highest value.

3.5. Decision Analysis

After characterizing the best predictive algorithm, we developed a decision analytics model to answer the following two questions:

1) Under which conditions should we use ML?
2) What is the extent of the ML cost savings?

To this end, we conducted an economic evaluation regarding the cost saving of ML implementation, i.e., implementing interventions based on the ML predictions (we refer to this as *ML-based strategy*) compared with the random strategy—selecting patients randomly. In practice, healthcare decision-makers may randomly assign an intervention to a group of patients (Suresh 2011; Cummings et al. 2022). To incorporate this, we consider random strategies by changing the implementation randomization, denoted by r, from 0% to 100%. There are also two extreme strategies sometimes used in practice, namely, implementing interventions on everyone, assuming that all patients are positive cases (we refer to this as "*treat-all" strategy*), and implementing interventions on no one, assuming that all patients are negative cases (we refer to this as "*treat-none" or "do-nothing" strategy*) (Vickers and Elkin 2006). These two extreme implementation strategies are special cases of the random strategy; namely, $r = 0\%$ represents the "treat-none" strategy, and $r = 100\%$ represents the "treat-all" strategy.

To this end, we calculated the four possible classification results of our predictive algorithms—that is, true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN). These

classification results are a direct function of the probability threshold P_t ($0 \leq P_t \leq 1$) above which the algorithm classifies the case as positive (i.e., fragmented readmission); otherwise, the case is classified as negative (i.e., nonfragmented readmission). Note that P_t can also be seen as the probability/risk threshold above which the decision-maker (here, the healthcare manager or policymaker) would select a patient for intervention; hence it represents the decision-maker's risk attitude. Accordingly, a low P_t means the decision-maker is risk-averse because he/she triggers the intervention at a low threshold, and a high P_t proxies a risk-loving decision-maker (Vickers et al. 2019). Because of the technical and practical implications of this threshold, we evaluated the above four classification results under a full range of P_t. After calculating the total cost of implementing each strategy at each threshold P_t, we chose the one yielding minimum cost. Finally, to gain broader decision-making insights, we performed extensive analyses wrt key factors, such as the effectiveness rate of interventions (e), predictive performance, event (i.e., fragmented and nonfragmented readmission) costs, and event rate.

Estimations of Personalized Cost. Our decision analysis aimed to estimate the optimal range of intervention costs over which each intervention strategy is optimal. To this end, we first estimated the patient-level costs of fragmented and nonfragmented readmission from data. These costs include patient-level expenses incurred during hospitalization, which are computed based on the patient's intensity of service utilization and acuity (Tran et al. 2020). Accordingly, each patient visit in the data is assigned a resource intensity weight (RIW) score characterizing an average level of hospital resources (consisting of administration, staff, supplies, technology, and equipment) utilized by the patient during each hospitalization (Wodchis et al. 2013). We applied the Case Mix Group Plus approach to estimate the personalized costs (Wodchis et al. 2013). In doing so, a year-specific RIW was used to account for the historical changes in resource utilization costs (e.g.,

internal practices, technology, and labor). Then, we estimated the patient-specific costs for all patient visits using the corresponding year-specific cost per weighted case (CPWC), reflecting the average cost incurred for a standard hospitalization, i.e., $RIW = 1.0$ (Gupta et al. 2021; Wodchis et al. 2013). Finally, all costs were adjusted for inflation based on 2020 Canadian dollars (CAD) using the appropriate Consumer Price Index (Bank of Canada 2021).

Cost Calculations. To estimate the population-based costs of fragmented and nonfragmented readmission, we averaged the hospitalization costs for all patients readmitted to non-index and index hospitals, respectively, as follows:

$$\bar{C}_{nonfragmented} = \frac{1}{n_1}\sum_{i \in Nonfragmented}^{n_1}\sum_{y=2004}^{2017}\left(RIW_{iy}^{Nonfragmented} \times CPWC_y\right), \quad (1)$$

$$\bar{C}_{fragmented} = \frac{1}{n_2}\sum_{i \in Fragmented}^{n_2}\sum_{y=2004}^{2017}\left(RIW_{iy}^{Fragmented} \times CPWC_y\right), \quad (2)$$

where $RIW_{iy}^{Nonfragmented}$ is the RIW for patient i at year y if readmitted to an index hospital, $RIW_{iy}^{Fragmented}$ is the RIW for patient i at year y if readmitted to a non-index hospital, and $CPWC_y$ is the year-specific cost per weighted case. For personalized analyses, the above calculations were stratified on the patient groups of interest. The details of cost calculations for different strategies are available in Appendix E1.

The total expected cost of each intervention strategy is the sum of the intervention expenses for n patients and the expected costs of readmissions (both fragmented and nonfragmented) after the intervention. Therefore, the expected total cost of the random intervention strategy depends on the cost of the intervention (applied to r% of patients), the total cost of the fragmented readmission (incurred because the intervention did not work), and the total cost of nonfragmented readmission. The expected total cost of the "random intervention" strategy can be calculated as follows:

$$TC_{random} = CI\ (n_r) + \bar{C}_{Fragmented}(n_2 - n_{2r}.e) + \bar{C}_{Nonfragmented}(n1 + n2_r.e), \quad (3)$$

where CI is the intervention cost; e refers to the intervention effectiveness rate—that is, how effective the intervention is in avoiding fragmented readmission ($0 \leq e \leq 1$)— n_r refers to the total number of randomly selected samples; $n2$ and $n2_r$ are, respectively, the number of patients with fragmented readmission estimated from the total data and $r\%$ randomly selected sample.

The expected total cost of the ML-based intervention strategy is the total cost of the intervention applied only to those who were predicted as "positive cases" by ML (i.e., TP+FP), plus the total costs of fragmented readmissions (incurred because the intervention did not work) and the total cost of nonfragmented readmission. In this strategy, even if the intervention is 100% effective, we still have a cost of fragmented readmission because of ML prediction errors. Please note that the fragmented readmission cost for this strategy also depends on the number of FN cases—that is, those who were falsely predicted negative by ML and hence, have no opportunity to receive the intervention. The expected total cost of the ML-based intervention strategy can be calculated as follows:

$$TC_{ML} = CI(TP + FP) + \bar{C}_{Fragmented}(TP(1 - e) + FN) + \bar{C}_{Nonfragmented}(TP.e + FP + TN). \quad (4)$$

Absolute and Relative Cost Savings. The expected absolute cost savings of the ML-based strategy can be calculated as follows:

$$\Delta_{random} = TC_{random} - TC_{ML}, \quad (5)$$

where Δ_{random} reflects total cost savings from utilizing the ML-based strategy compared with the "random intervention" strategy. To convert the dollar value into a percentage of cost saving, the relative cost savings can be calculated as follows:

$$\delta_{random} = \frac{TC_{random} - TC_{ML}}{TC_{random}} \times 100. \quad (6)$$

Intervention Cost Optimization. For our decision analysis, we take that intervention cost is an exogenous variable and identify the range of intervention costs over which each intervention strategy is optimal. The optimal value of intervention cost can be calculated by equating the associated absolute cost-saving functions with zero, as follows:

$$CI^* = \frac{\left(\left((1-e)TP+FN\right)-n_2+e.n_{2r}\right)\bar{C}_{Fragmented}+\left((TP.e+FP+TN)-(n_1+n_{2r}.e)\right)\bar{C}_{Nonfragmented}}{n_r-(TP+FP)} \quad (7)$$

For richer insights, we conducted extensive analyses regarding the optimal range of intervention costs wrt important factors, including probability threshold (P_t), intervention effectiveness rate (e), ML's predictive performance, and the costs of fragmented and nonfragmented readmission.

3.6. Fairness Analysis

Existing approaches to algorithmic fairness typically follow a two-step process. First, a formal notion/definition of fairness is defined; then, a decision rule is developed to satisfy this measure (Corbett-Davies et al. 2017). In the last few years, more than 20 different notions of fairness have been proposed for algorithmic problems (Verma and Rubin 2018). These notions are classified in terms of two doctrines of fairness, namely, 'treatment parity' and 'impact parity,' which aim to address different causes of non-discriminatory decisions (Fu et al. 2021). Whereas treatment parity involves treating equal people equally regardless of their membership in a protected class (i.e., equality), impact parity addresses outcome parity, and it reflects an idea of equal "opportunity" to access a benefit for eligible people (i.e., outcome equality/equity) (Fu et al. 2020, 2021). Most literature on algorithmic fairness has focused on its statistical definitions. They approximate parity in terms of some model performance measures (e.g., positive classification rate or statistical parity,

FP rate [FPR] and FN rate [FNR] or equalized odds, PPV or predictive parity, etc.) against or in favor of protected attributes (e.g., race/ethnicity, sex, income, etc.) (Chouldechova and Roth 2018).

There are several statistical definitions of fairness, which can be categorized into three groups: (i) definitions based on predicted outcomes, (ii) definitions based on predicted and actual outcomes, and (iii) definitions based on predicted probabilities and actual outcomes (see Verma and Rubin 2018). Our study considered "statistical parity" from the first category, "predictive parity," "predictive equality," "equal opportunity," "accuracy parity," "FNR parity," "FPR parity," "negative predictive value (NPV) parity," and "receiver operating characteristic (ROC) AUC parity" from the second category; and "calibration" from the last category as the main fairness concerns because of their higher prevalence in the context of healthcare and their potentially harmful impact on patients if left unresolved. These measures ensure certain perspectives of impact parity and are widely discussed and used in the literature (Fu et al. 2020). We also considered a newly developed comprehensive measure from the second category— that is, Matthews correlation coefficient comparison (MCC) parity (Chicco et al. 2021; Verma and Rubin 2018). MCC summarizes both positive and negative cases with the same weight of importance into a single informative metric. A high value of MCC means that the predictive model correctly predicted most of the positive and negative cases, and most of these positive and negative predictions are correct (Chicco et al. 2021; Chicco and Jurman 2020). For fairness analysis, we used the "*fairness"* package in R 3.6.0. The definitions and formulas for each notion are summarized in Appendix F (Table F1).

While we focused on the "impact parity" for the ML-based strategy, the random strategy ensures "treatment parity" because each patient has an equal chance of receiving the intervention

(Cummings et al. 2022; Suresh 2011). Random assignment prevents selection bias by distributing the patients' characteristics (Akobeng 2005); however, this may contribute to inequitable and avoidable differences in population health outcomes (Qi et al. 2021). Consequently, to provide insights into the effect of these two major theories of fairness on decision-making, we compared our ML-based strategy with the random strategy. Our analysis considered sex, marginality (ethnicity), and residency type (rural/urban) as the three legally recognized protected groups.

It is not the case that we cannot do anything about bias. To mitigate or remove algorithmic bias, the three following main approaches can be taken: (i) preprocessing (transforming the imbalance data of the protected groups), (ii) intermediate processing (model improvement during training, such as adding fairness constraints while ML training), and (iii) postprocessing (adjusting final prediction results, e.g., group-specific threshold setting) (Fu et al. 2020; Rajkomar et al. 2018). We corrected algorithmic bias by processing the final prediction results rather than changing the ML design and raw data. Most postprocessing approaches involve setting a group-specific threshold and assigning different decision thresholds for the protected and nonprotected groups as corrections for disparate impact (Fu et al. 2021). For example, for the sake of "equal opportunity" among sexes, one can set two different decision thresholds (P_t) for women and men such that the fraction of protected members (i.e., women) and nonprotected members (i.e., men) that qualify for receiving preventive intervention is the same (Hardt et al. 2016). Although this technique is simple and intuitive (Fu et al. 2021), multiple thresholds would be required to satisfy any notion of fairness (e.g., predictive parity, predictive equality, equal opportunity, and equalized odds) because choosing one threshold could satisfy one definition but lead to an unfair solution for the others. However, policymakers would prefer to decide according to one decision rule (Corbett-Davies et al. 2017).

4. Results

4.1. Descriptive Results

Descriptive analyses were performed for the nonfragmented (NF) versus fragmented (F) readmissions. The results of the summary statistics are provided in Table B1 (Appendix B). The study included a total of 953,818 observations of patients aged 65–102 readmitted within 30 days. Among them, 61% were readmitted to the index hospital, and 39% were readmitted to the non-index hospital. In the analytic cohort, the mean age (standard deviation) of the patients was 75.1 (7.6) years; 29.5% were older than 80 years, and 51.5% were female. Patients experiencing a fragmented readmission—that is, patients readmitted to a non-index hospital—were more likely to reside in rural areas than those with nonfragmented readmission (F: 28.29% vs. NF: 15.84%, $p < 0.0001$); more likely to have a routine discharge to home (F: 71.98% vs. NF: 51.43%, $p < 0.0001$); more often admitted with a primary diagnosis of acute myocardial infarction (F: 10.02% vs. NF: 3.25%, $p < 0.0001$); and less often admitted with a respiratory disease, such as pneumonia (F: 1.70% vs. NF: 3.09%, $p < 0.0001$) and COPD (F: 2.93% vs. NF: 8.03%, $p < 0.0001$) in their index admission. The top causes for readmission to a non-index hospital were angina pectoris, chronic ischemic heart disease, acute myocardial infarction, fracture of the femur, and convalescence following surgery, whereas the prevalent causes of readmission to an index hospital were COPD, urinary tract infection, complications after surgery, and intestinal obstruction. In addition, patients whose readmission was fragmented were more likely to visit a special care unit during hospitalization (F: 28.14% vs. NF: 13.91%, $p < 0.0001$), with a LOS of up to 3 days (F: 12.58% vs. NF: 6.71%, $p < 0.0001$). Overall, 7.9% of patients died in the hospital after their 30-day readmission. The average resource intensity weights, representing readmission charges, were also higher among the fragmented readmission group than its nonfragmented counterpart (F: 2.68

vs. NF: 1.95, $p < 0.0001$). Regarding the average readmission LOS, statistically significant differences were seen for nonfragmented versus fragmented readmission groups (F: 10.67 vs. NF: 11.33, $p < 0.0001$). The results of the association analysis (Appendix A) also showed that the fragmented readmission was strongly associated with an increase in both resource utilization outcomes (Odds Ratio $(OR)^{30d}_{LOS}$ = 1.12, 95% Confidence Interval (CI) = 1.11–1.13, p-value <0.001; OR^{30d}_{RIW} = 1.42, 95% CI = 1.40–1.44, $p < 0.001$), but not with increased in-hospital mortality following readmission ($OR^{30d}_{mortality}$ = 0.98, 95% CI = 0.96–1.00, p = 0.007).

4.2. ML Predictive Performance

Table 1 presents various predictive performance measures that were examined on the test set for all utilized ML algorithms, along with different resampling approaches.

Table 1. Out-of-sample ML predictive performance with different resampling approaches

Algorithm	Resampling technique	AUC	Recall/ Sensitivity	Specificity	Precision	F1	Accuracy
NB	**None**	0.68	0.46	0.78	0.70	0.74	0.66
	Up-sample	0.68	0.51	0.76	0.71	0.74	0.66
	Down-sample	0.68	0.51	0.76	0.71	0.74	0.66
LR	**None**	0.73	0.48	0.86	0.72	0.79	0.71
	Up-sample	0.73	**0.68**	0.67	0.77	0.72	0.68
	Down-sample	0.73	**0.68**	0.67	0.77	0.72	0.67
CART	**None**	0.73	0.48	0.87	0.73	0.79	0.72
	Up-sample	0.73	0.63	0.75	0.76	0.75	0.70
	Down-sample	0.73	0.61	0.75	0.76	0.77	0.70
XGB	**None**	**0.75**	0.49	**0.88**	0.73	**0.80**	**0.73**
	Up-sample	0.74	0.62	0.75	0.76	0.75	0.70
	Down-sample	**0.75**	0.61	0.77	0.76	0.76	0.71
RF	**None**	**0.75**	0.49	**0.88**	0.73	**0.80**	**0.73**
	Up-sample	**0.75**	0.62	0.77	0.76	0.76	0.71
	Down-sample	**0.75**	0.63	0.75	**0.77**	0.77	0.70
ADA	**None**	**0.75**	0.50	0.86	0.73	**0.80**	**0.73**
	Up-sample	**0.75**	0.65	0.72	**0.77**	0.74	0.69
	Down-sample	**0.75**	0.65	0.72	**0.77**	0.74	0.69
Stacked Ensemble[§]	**None**	**0.75**	0.49	0.87	0.73	**0.80**	**0.73**
	Up-sample	0.74	0.59	0.78	0.75	0.77	0.71
	Down-sample	**0.75**	0.64	0.74	**0.77**	0.75	0.70

Note: The boldfaced numbers indicate the highest value of a predictive performance indicator
[§] The base learners are XGB, RF, ADA, and the super learner is LR

These results indicate very good prediction generalization without overfitting the unseen data. Table C1 (Appendix C) presents the estimated predictive performance measures for 5-fold cross-validation of the ML algorithms to predict fragmented readmission. The details of hyperparameter tuning are available in Appendix C (Table C2). Regarding predictive performance measures, particularly AUC and F1-measure, NB performed the worst, while RF, XGB, and ADA exhibited the best performance. Although the boosting-based algorithms (i.e., XGB and ADA) and the bagging-based algorithm (i.e., RF) yielded similar predictive performance, they exhibited different performance variance and runtime results. The XGB ran faster compared with the RF and ADA, with less performance variance. LR performance measures were modestly lower than those of the winner algorithms (i.e., XGB, ADA, and RF). The ROC curves for all constructed models are presented in Figure 2 to understand the effects of decision thresholds on model performance. The results denote that the resampling approach did not significantly affect AUC.

To gain a deeper understanding of the ML models' performance based on the combination of predictive performance measures, we conducted a Pareto analysis. To this end, we analyzed the Pareto optimality point and frontier line to compare the predictive models based on multiple performance measures and identify the nondominated solutions (results in Appendix C3). First, we found ten nondominated ML algorithms based on the six performance measures (i.e., AUC, F1, sensitivity, specificity, precision, and accuracy) (Table C3). Next, for simplicity and ease of visual illustrations, we reduced the six-dimensional Pareto analysis to a three-dimensional analysis to identify nondominated models in terms of three performance measures that cover all six measures. These were AUC, F1, and accuracy because AUC is a function of sensitivity and specificity and F1 is a function of precision and recall. The remaining dimension was accuracy, which we considered along with AUC and F1. The results (Figure C1) indicate that the

nondominated ML models were ADA, XGB, and RF. Interestingly, this result did not violate the six-dimensional Pareto analysis.

Finally, to assess the agreement between the predicted and estimated risk of the outcome, we examined calibration power for all constructed ML algorithms. The results (Appendix C; Figure C2) indicate that XGB and RF provided the highest calibration, while NB and ADA were poorly calibrated. These results are consistent with the binary prediction performance of the algorithms, providing even greater support for the superiority of XGB and RF. We considered XGB the winner for our remaining analysis because this algorithm runs significantly faster than RF.

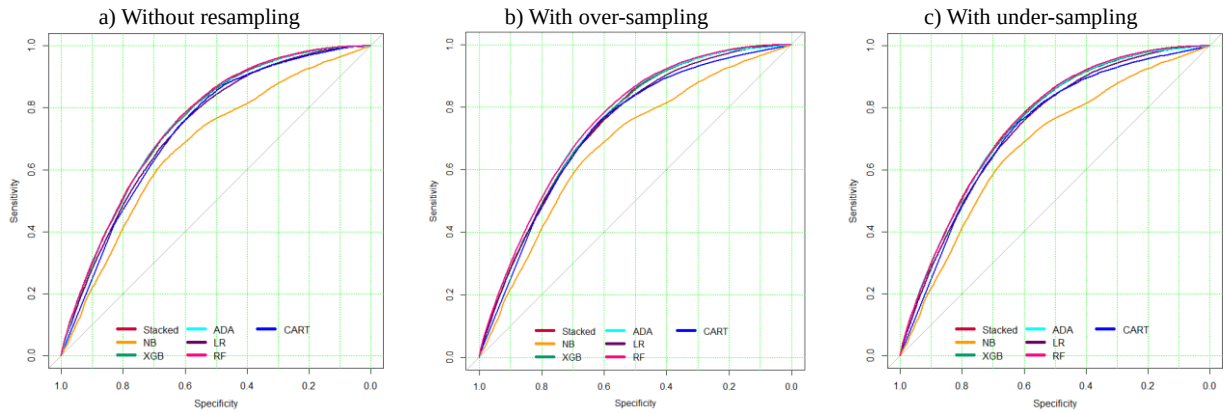


Figure 2. ROC curve for the constructed predictive models with different resampling approaches.

Note: XGB: extreme gradient boosting, LR: logistic regression, CART: classification and regression tree, NB: Naïve Bayes, RF: random forest, ADA: adaptive boosting

The results (Figure C1) indicate that the nondominated ML models were ADA, XGB, and RF. Interestingly, this result was not violated by the six-dimensional Pareto analysis. To assess the agreement between the predicted and estimated risk probability of the outcome, we examined the calibration curve for all constructed ML algorithms. The results (Appendix C; Figure C2) indicate that XGB and RF provided the highest calibration, while NB and ADA were poorly calibrated. These results are consistent with the discrimination power of the algorithms, providing even greater support for the superiority of bagging and boosting winner algorithms (XGB and RF). We

considered XGB to be a winner for our remaining analysis because this algorithm runs significantly faster than RF.

4.3. Results of Decision Analysis

Our cost estimations (details in section 3.8) indicate that the average fragmented readmission cost per patient was $11,344, and the nonfragmented readmission cost per patient was $9,448. Considering our best predictions obtained from XGB, we calculated the relative cost savings of random strategies vs. ML-based strategy for a range of effectiveness rates and threshold probabilities to identify the range of intervention costs over each strategy is optimal. The general shape of these results is depicted in Figure 3, which is interesting for its threshold nature, facilitating its implementation in practice. Accordingly, the results present three optimality areas for (i) greedy randomization (GR), where the level of randomization is low, (ii) aggressive randomization (AR), where the level of randomization is high, (iii) an ML-based strategy. The results can be interpreted as follows: i) A low level of randomization ($r \leq r_1$), i.e., GR, is optimal only if the intervention cost is more expensive than a certain threshold ($CI \geq C_1$). ii) A high level of randomization ($r \geq r_2$), i.e., AR, is optimal only if the intervention cost is less expensive than a certain threshold ($CI \leq C_2$). iii) In all other cases, the ML-based strategy is the optimal implementation strategy. More specifically, for a certain range of randomization ($r_1 < r < r_2$) and a certain range of intervention costs ($C_1 < CI < C_2$), the ML-based strategy is always the implementation strategy.

The four threshold values (hence the optimality areas) depend on other key parameters, such as the decision-makers' risk attitude P_t and the intervention effectiveness e. To investigate this, Figure 4 shows the results for varying e while keeping P_t fixed at 0.5 (indicating a risk-neutral decision-maker).

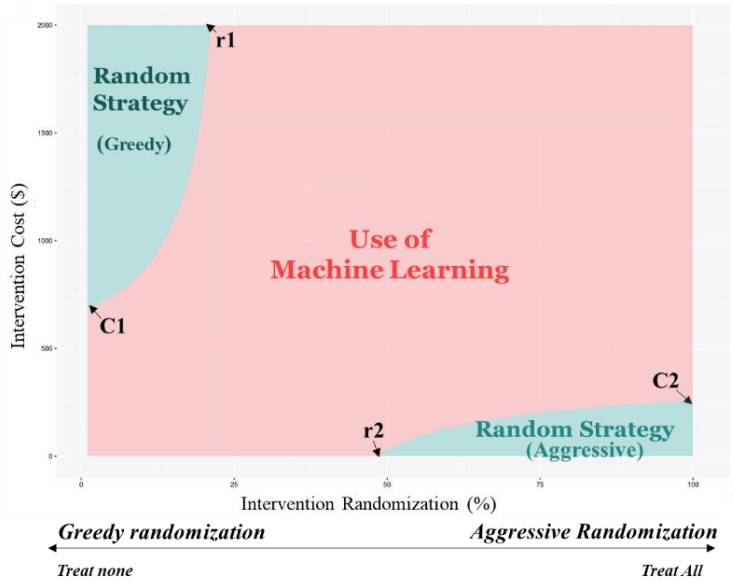


Figure 3. Optimality areas of ML wrt the random strategies

We observe that when the intervention effectiveness improves (from 0 to 1), the GR area shrinks in favor of the AR area's expansion. The results are intuitive because, when an intervention is less effective (for example, look at the panel of $e = 0 - 0.1$), it must be used more cautiously, i.e., GR strategy and there is little chance for an aggressive strategy (i.e., AR) to be successful (unless the intervention is excessively cheap). Interestingly though, the ML-based implementation of such a low-effective intervention is preferred if the intended level of random implementation increases (beyond GR). In contrast, when an intervention becomes more effective (for example, look at the panel of $e = 1$), it can be used more generously (i.e., AR), particularly if the intervention is cheap, and a cautious random implementation of it becomes less reasonable. Again, under this scenario, for a wide range of parameters, a smart prediction-based implementation of the intervention (i.e., using ML) must be the preferred choice.

Next, we reconducted the analysis by varying the probability threshold or decision-makers' risk attitude (P_t) while fixing the intervention effectiveness (e) at 0.5. Figure 5 illustrates the optimality

areas for this analysis. Interestingly, the impact of changing P_t on the optimality areas is quite different than that of e.

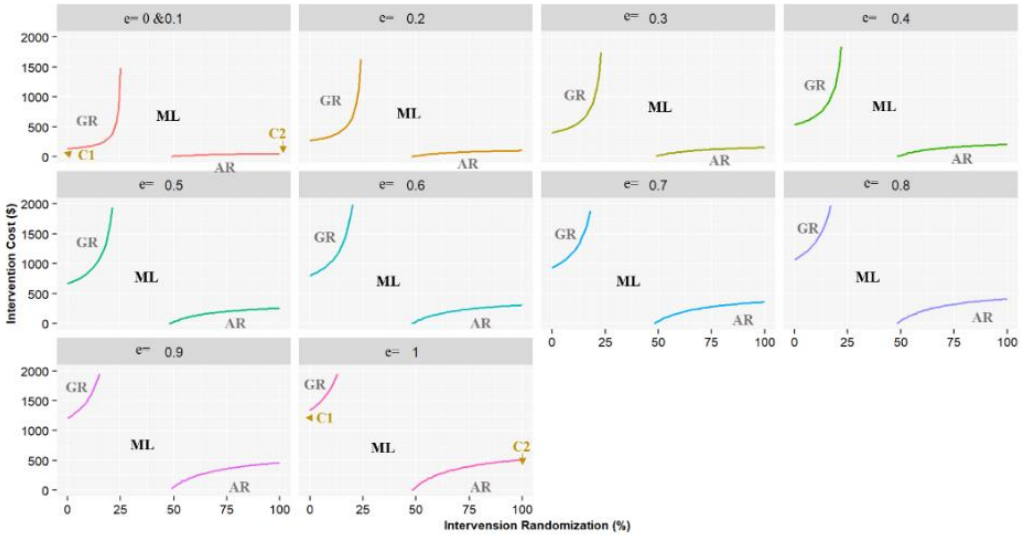


Figure 4. Optimality areas when intervention effectiveness changes ($P_t = 0.5$)

The first interesting observation is that when the risk threshold is extremely low (i.e., $P_t \leq 0.1$), representing a very conservative approach, ML coincides with the treat-all strategy because, under this low-risk threshold, ML predicts everything as positive, hence recommending intervention for everyone. In this case, all randomization levels are preferred to ML (which is the same as $r = 1$ randomization) if intervention is more expensive than a certain value (here $365); otherwise, ML must be preferred. For extremely high-risk thresholds (i.e., $P_t \geq 0.9$), ML coincides with the treat-none strategy because it predicts everything as negative, hence refuting intervention for everyone. In this case, all randomization levels are preferred to ML (which is the same as $r = 0$ randomization) if intervention is less expensive than a certain value (here $365); otherwise, ML must be preferred. In all intermediary cases, as the risk threshold increases (from 0.2 to 0.8), the GR's optimality areas shrink in favor of expanded AR's optimality area.

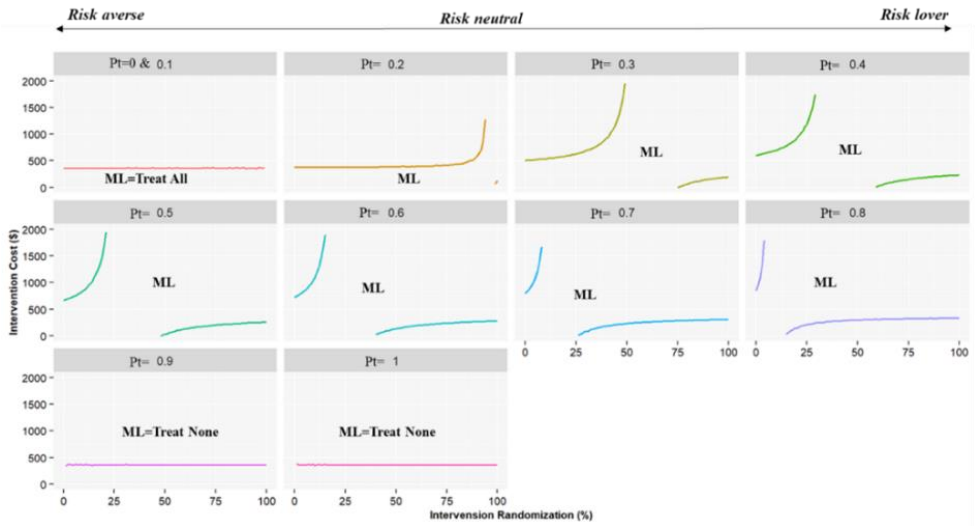


Figure 5. Optimality areas when risk threshold changes ($e = 0.5$)

Figure 6 presents the optimality areas when changing the cost of the fragmented readmission (fixing $e = 0.5, P_t = 0.5$). In this analysis, we kept the cost of nonfragmented readmission unchanged. As the results show, the GR optimality area shrinks when the difference between fragmented and nonfragmented readmission increases (from the green to the pink line). This means that when fragmented readmission becomes costlier, the GR's optimality area shrinks in favor of the ML optimality area, while AR's optimality area expands against the ML optimality area. We have opposite results when the difference between fragmented and nonfragmented readmission decreases (from green to the blue line).

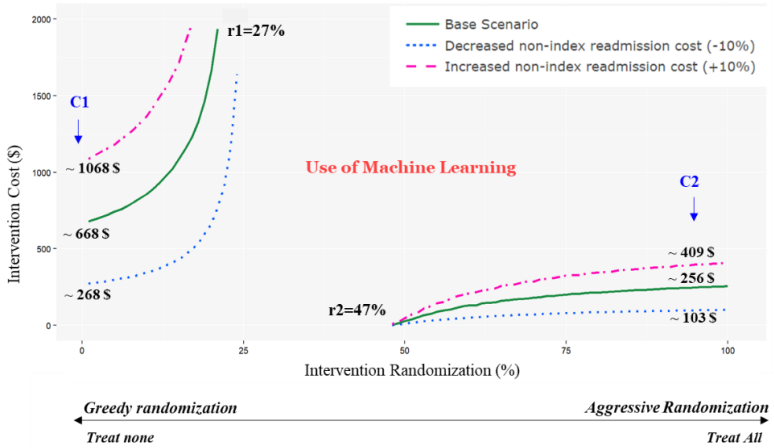


Figure 6. Optimality area when the cost difference between fragmented and nonfragmented readmission changes ($e = 0.5, P_t = 0.5$)

We next investigated how the optimality areas change with the ML predictive performance. From Figure 7, we can see that the ML's specificity impacts the GR area, whereas the ML's sensitivity affects the AR area. Recall that sensitivity refers to the model's ability to correctly detect patients with fragmented readmission who have the condition (i.e., $TP\ /\ (TP + FN)$). It shows how interventions are correctly assigned to eligible patients. It is positively associated with TP ($TP = n_2 \times sensitvity$) and negatively associated with FN ($FN = n_2 \times (1 - sensitvity)$) (Weiner et al. 2018). Conversely, specificity relates to the ML's ability to correctly disqualify patients with nonfragmented readmission for the intervention. It is positively associated with TN ($TN = n_1 \times specificity$) and negatively associated with FP ($FP = n_1 \times (1 - specificity)$) (Weiner et al. 2018).

The ML strategy recommends imposing the intervention only on the positive cases (i.e., TP+FP); however, the AR strategy recommends intervention for more than this rate (i.e., $\geq (\mathrm{TP} + \mathrm{FP})/n$). Intuitively, the AR strategy offers a greater chance for patients who have mistakenly been predicted as a negative case by ML (i.e., FN) to receive an intervention. Hence, more returns can be achieved from the AR strategy compared with the ML strategy when sensitivity decreases. However, this is reasonable for inexpensive interventions ($C_I \leq C_2$) because of the tradeoff between the intervention cost and savings from true preventions (fragmented readmissions). Hence, for any $C_I > C_2$, the savings of the random strategy (compared to ML) will become negative.

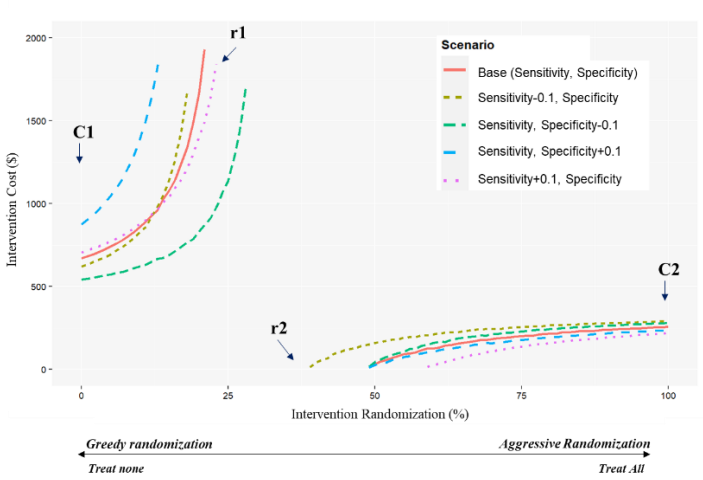


Figure 7. Optimality area of strategies: sensitivity analysis on the model performance ($e = 0.5, P_t = 0.5$)

Finally, thanks to the ML ability to provide person-level predictions and our personalized estimations of the costs, we were able to provide personalized recommendations. As mentioned before, CF is more common among medically complex patients, who are identified based on multimorbidity and frailty. Therefore, we conducted personalized decision analysis for four patient groups based on their frailty and the comorbidity burden. To simplify the analysis, we fixed the value of effectiveness and threshold at 0.5. Also, the event rates across the groups were roughly equal (37%). The results (Figure 8) indicate that the optimality areas are different for patient groups because of their differences in ML performance and differential cost of readmissions. The results show that ML is most preferred to GR for patients with high frailty and comorbidity. The model performance (particularly sensitivity) was not promising for this group (sensitivity=0.36; specificity=0.89). However, the differential cost of readmissions was much higher for this group than for other groups ($8,425). Hence, this scenario investigated the concurrent impact of a costly event with model performance (here, low sensitivity and high specificity). As expected, the higher differential cost of readmissions for the most complex patients increased the ML strategy's utility wrt GR. This effect was magnified by the high value of specificity for this group. Conversely, the costly event and a model with low sensitivity synergistically increased the optimality area of the

AR. This result is intuitive because when an event is costly, and the ML model does not perform well in predicting positive cases (low sensitivity), an aggressive random strategy achieves more returns, gaining greater favorability. Details are available in Appendix E (Table E1).

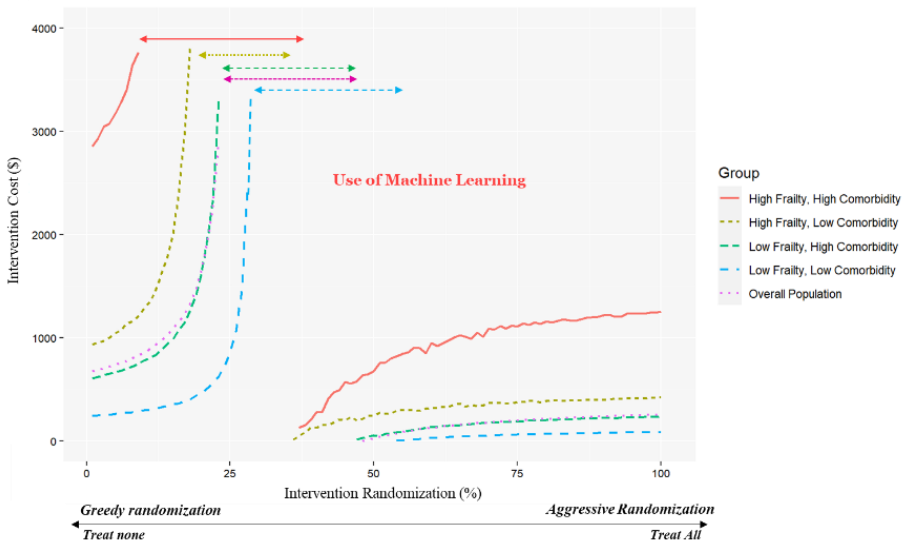


Figure 8. Personalized intervention cost recommendations for patient complexity groups ($e = 0.5, P_t = 0.5$)

4.4.Results of Fairness Analysis

Evaluating Fairness among Protected Groups. Table 2 summarizes various measures of group-specific fairness for sex, marginality, and residency type. We used the "80 percent"/ "four-fifth" rule to assess the fairness of our ML model. This rule—suggested by the US Equal Employment Opportunity Commission— requires a maximum 20% deviation of the selection rate among subgroups depending on the situation (Barocas et al. 2017; Qi et al. 2021). Given the 80% rule in disparate impact law, our model satisfies all fairness definitions for sex and marginalization as the fairness scores are greater than 0.8. However, it does not perform well in satisfying several notions of fairness for residency, including statistical parity, equal opportunity, predictive equality, FNR, and FPR parity.

Table 2. Comparison of the fairness scores of the ML predictions among different protected subpopulations

Fairness Definition	Sex			Residency			Marginalization		
	Male	Female	$\frac{Female}{Male}$	Rural	Urban	$\frac{Rural}{Urban}$	Low-Moderate	High	$\frac{Low-Mod.}{High}$
Statistical parity (SP)	0.26	0.27	**1.03**	0.55	0.19	**2.89**	0.26	0.26	**1.00**
Equal opportunity (EO)	0.48	0.48	**1.00**	0.74	0.38	**1.95**	0.48	0.48	**1.00**
Predictive parity (PP)	0.70	0.71	**1.01**	0.72	0.69	**0.97**	0.71	0.70	**1.00**
Predictive equality (PE)	0.87	0.88	**1.01**	0.68	0.91	**0.75**	0.87	0.87	**1.00**
Accuracy parity (ACC)	0.72	0.73	**1.00**	0.71	0.73	**0.97**	0.72	0.72	**1.00**
FNR parity	0.52	0.52	**1.00**	0.25	0.62	**0.41**	0.52	0.52	**1.00**
FPR parity	0.13	0.12	**0.94**	0.33	0.09	**3.70**	0.13	0.13	**1.00**
NPV parity	0.73	0.73	**1.00**	0.71	0.73	**1.02**	0.73	0.73	**1.00**
AUC parity	0.75	0.75	**1.00**	0.78	0.72	**1.08**	0.75	0.75	**1.00**
MCC parity	0.39	0.40	**1.02**	0.42	0.35	**1.2**	0.39	0.39	**1.00**

ML-based models satisfy fairness in terms of *statistical parity* if both protected and regular groups have an equal probability of being detected (through predictions) as fragmented readmission (hence receiving a preventive intervention). The results showed that the proportion of predicted positive cases for rural and urban residents were 0.55 and 0.19, respectively ($SP(rural/urban) = 2.89$). Although this result favors the protected group, we can conclude that the ML does not perform well in satisfying this notion of fairness for residency. Given the value of *predictive equality*, the ML performed poorly for residency (rural: 0.68 vs. urban: 0.91). We observed patterns of disparity among patients from rural and urban areas. Rural residents were at a higher risk of being falsely flagged as positive cases (i.e., with fragmented readmission) and of receiving unnecessary preventive interventions ($FPR(rural/urban) = 3.70$). The FP results, which can also lead to stress or overtreatment (Keskinocak and Savva 2020), might raise alert fatigue over time and can desensitize care providers so that they override critical cases as a result of mistrust of the system's prediction (Poly et al. 2020). Given the *equal opportunity* value for rural patients (0.74) and the urban population (0.38), our model did not perform well in terms of satisfying this definition of fairness for this underserved group ($EO(rural/urban) = 1.95$). Our results (Table 2) indicate that rural groups are at a higher risk of being falsely predicted to have unfragmented readmission and thus missing out on preventive interventions ($FNR(rural/urban) = 0.41$). However, given the MCC parity and the four-fifth rule, our model satisfies all fairness definitions

for the three protected groups. We also investigated the "calibration" definition of algorithmic fairness that compared the predicted probabilities for both protected and nonprotected groups within the full range of probability thresholds ($0 \leq P_t \leq 1$). The results, which are available in Appendix F (Table F4), indicated that our ML-based model was well-calibrated for all threshold values wrt sex and marginalization but only for $P_t > 0.5$ wrt residency. In the next section, we discuss the correction method to minimize the issue of algorithmic bias for the residency attribute.

It should be noted that compared with the ML-based strategy, which focuses on impact parity (as discussed above), the random strategy satisfies fairness through unawareness/blindness; that is, the sensitive attributes are not explicitly used in patient selection for the intervention. Although this notion of fairness, which represents equal treatment, is intuitive and straightforward, it often leads to different outcomes across the protected groups; hence, it may not be sufficient to ensure impact parity (Fu et al. 2020).

Evaluating Fairness among Intersectional Groups. We also assessed the fairness among intersectional groups to the extent of unfairness for the intersections of the attributes will be, such as highly marginalized females. We found that in an intersectional group, the extent of fairness and unfairness remains the same. For instance, the results on the equal opportunity (Figure 9)—the most common fairness notion— showed that rural females have higher EO than urban females (0.74 vs. 0.36). Given the similar EO for sex ($EO_{female\ vs.\ male}$: $0.48\ vs.\ 0.48$), the difference was exactly reflected the unfairness for the residency attribute ($EO_{rural\ vs.\ urban}$: $0.74\ vs.\ 0.38$). The critical social/policymaking implication of this is that we do not need to focus on those attributes which are fair (e.g., sex); instead, we only focus on the unfair attributes. Because of the intersections, the extent of unfairness will not change. The results can be interpreted the same for the other fairness notions, such as predictive equality (Figure F2).

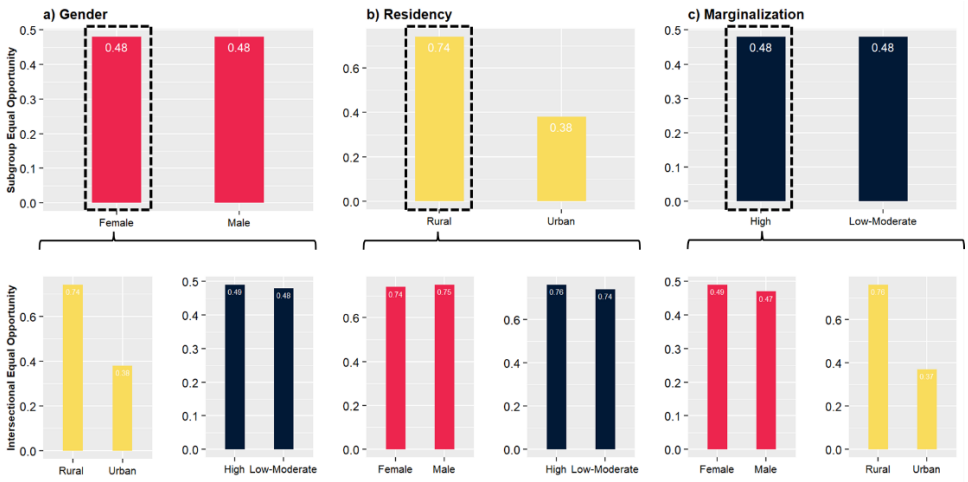


Figure 9. Analysis of equal opportunity across subgroups and intersectional groups

Correction Approaches to Mitigate Decision Bias. The previous section suggested that our ML predictions are fair for two sensitive attributes—namely, sex and marginalization. However, our model performed poorly in satisfying some important fairness measures for the residency attribute. Therefore, we adopted a postprocessing correction approach to adjust the predictions toward fair recommendations for this protected group. To this end, we first investigated the feasibility of setting group-specific thresholds for each notion of fairness. This approach focuses on the two most popular notions of fairness, which are equal opportunity and predictive equality (Fu et al. 2020). The analysis for satisfying equal opportunity for the residency attribute led to thresholds of 0.4 for urban residents and 0.7 for rural residents. To find the thresholds, we set P_t=0.5 and decrease (increase) the threshold until both group members receive a similar fairness score. These are the thresholds at which around 50% of both group members would be qualified to receive the intervention. This resulted in fairness in terms of equal opportunity ($equal\ opportunity_{rural} = 0.48; equal\ opportunity_{urban} = 0.49$). Interestingly, the predictive equality was also satisfied with the same threshold ($predictive\ equality_{rural} = 0.88$; $predictive\ equality_{urban} = 0.86$). More details are available in Appendix F (Tables F2–F3). Figures 11a, b display how ML

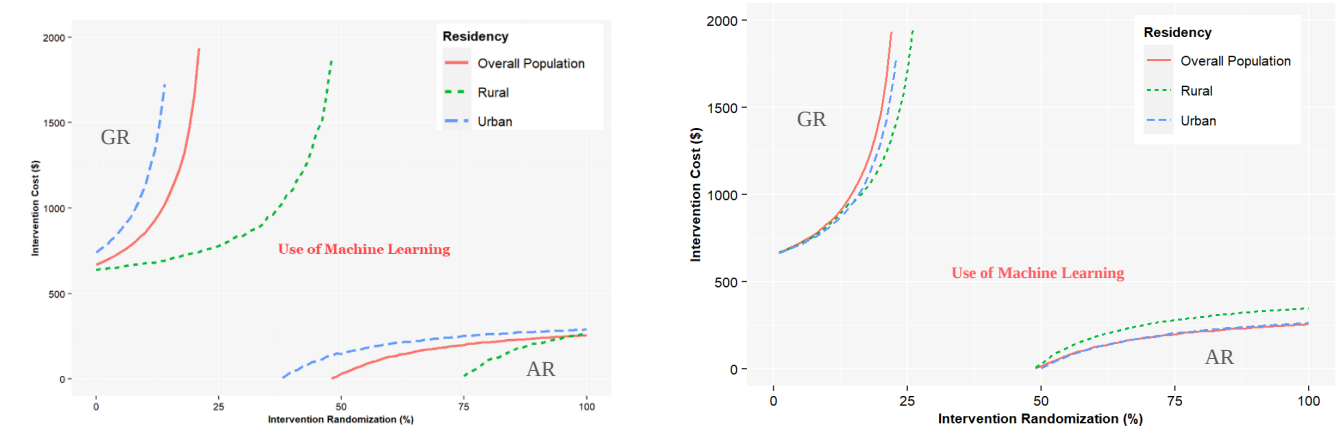
optimality areas were modified for the residency members (e.g., rural vs. urban) after applying the post-correction procedure. The greater similarity in the optimality areas was achieved using a group-specific threshold for rural and urban groups. However, a slight difference could be observed in the optimal intervention costs. This difference is a natural and necessary consequence of enforcing fairness into our framework, leading to fairness according to the *need-based philosophy* discussed below.

It should be noted that the two notions of fairness described above were applied only to the predictive analytics part, not necessarily to the entire decision-making framework (i.e., predictive-prescriptive). To resolve this issue, we need to introduce two other criteria, i.e., event rate (i.e., the prevalence of fragmented readmission: $\frac{positive\ cases}{total\ cases}$) and differential cost of the fragmented readmissions, in addition to the predictive fairness notions discussed above. We included these two criteria (i.e., cost and event rate considerations) in our decision-making framework for the following reasons: according to the need-based resource allocations philosophy, for an equitable resource allocation, patients with greater (lesser) needs should receive more (less) resources (Anselmi et al. 2015). Evidence suggests that need-based resource allocations are more likely to enhance equity (Anselmi et al. 2015; Radinmanesh et al. 2021). A higher cost of fragmented readmission can be translated to higher resource utilization (because it is a direct function of RIW), which means that more services are needed. Moreover, patient groups with a higher risk of facing the event require intensified health services (Culyer 2007). Notably, the event rate (which concerns the actual positive cases) is different from predictive fairness notions, such as "equal opportunity" (which concerns predicted risk and has been resolved using the post-correction method delineated above). Figure F1 (Appendix F) shows the difference between the event rate and equal opportunity for the residency attribute.

We start with the residency attribute to discuss the reasons for the differences in the optimality area (after fixing the algorithmic bias using the post-correction procedure) and its association with fair decision-making from a need-based philosophy perspective. The cost of fragmented readmission is $1,774 higher than that of nonfragmented readmission among rural residents, while this difference is $2,159 among urban residents (Table F7). In contrast, event rates are higher among rural residents than they are in urban groups (53% vs. 34%). Following the need-based resource allocations philosophy to attain fair distribution of resources, our system suggests a higher intervention threshold for rural patients ($349) than for urban patients ($265) for the strategy AR (i.e., C_2 in Figure 4). Although the cost difference was higher among urban populations than rural ones ($385), favoring a higher intervention cost for the urban group (Figure 10b), the impact of a higher event rate among rural patients dominated the effect of costs, contributing to a decision in favor of rural residents. In other words, the suggested approach balances the tradeoff between two impacts (cost vs. event rate). More specifically, had we not considered the differential cost of fragmented readmission among different groups, we could have implemented fairness from a predictive (i.e., risk) perspective only but violated another key dimension of fairness which accounts for the difference among different groups based on their needs and complexities. For example, two different groups may have the same risk (before or after enforcing risk-based fairness) but not the same need, cost, or complexity. The equity-based notion of fairness mandates treating these two groups differently. In our framework, we control both. More specifically, the first one is achieved through our efforts to enforce fairness on the ML side, and the second one is achieved through our personalized decision-making framework. The results are aligned with our sensitivity analysis regarding the event rate (Appendix E, Figure E2), which ascertains that a higher event rate expands the AR optimality area.

a) Optimality area of residency before correction

b) Optimality area of residency after correction



a) ML optimality area of marginalization

b) ML optimality area of different sexes

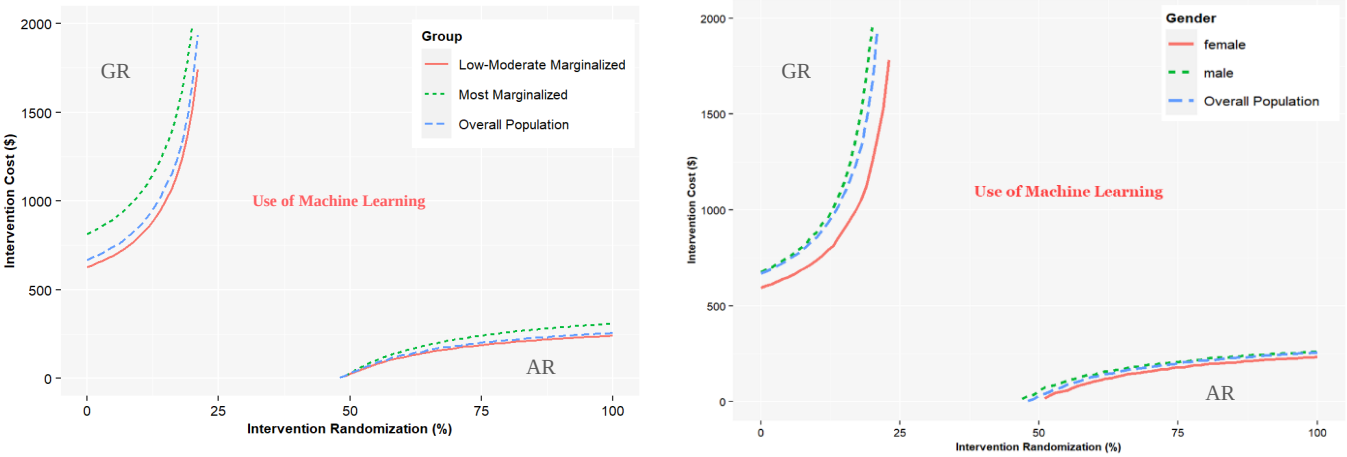


Figure 10. Comparing ML optimality areas for the protected groups

Regarding the other two sensitive attributes—sex and marginalization—we showed that our framework satisfied the predictive notions of fairness for these two protected groups in the previous section. Next, we investigated the impact of two other criteria for fairness (i.e., cost and event rate) in our predictive–prescriptive framework. Our results (Table F5) indicate that the cost difference between fragmented and nonfragmented readmission is greater among groups with high marginalization versus low to moderate marginalization ($2,302 vs. $1,786). Our decision-making framework reflects these differences in the suggested optimal intervention costs. Figure 10c shows that the optimal intervention costs are $310 versus $241 for groups with high versus low to moderate marginalization wrt the AR (i.e., C_2 in Figure 3). This suggests that higher resource

allocation ($) should be made to the group with high marginalization group compared with the group with low to moderate marginalization because the former group has higher needs. Because the rate of events is roughly similar among groups with low to moderate and high marginalization (0.37 vs. 0.39), we can conclude that the differences in the optimal intervention costs stem only from the cost difference. The same pattern can be seen by comparing the results of men versus women (Figure 10d). The cost differences are $1,665, and $1,973, respectively, for women and men (Table F6). The suggested optimal intervention costs are $233 for women and $261 for men. As for the marginalization attribute, no significant difference was found between the event rates of men versus women (0.38 vs. 0.40); hence, the difference in cost was the main contributor to this dissimilarity in the optimal intervention costs.

4.5.Real-World Application

Health policymakers aim to decide about the most promising preventive intervention to mitigate CF among the aging population under budgetary constraints; the key question we seek to address using our prediction-informed decision-making framework is the following: given a limited budget B and m intervention programs with different implementation costs (CI^j) and effectiveness rates (e^j), where $j = 1, \dots, m$, how do we choose the best course of action (the best prevention planning) to the best economical way (economic perspective)? As illustrated in the previous section, our proposed decision-making framework also reduces disparities among the protected groups (ethical perspective).

Transitional Care Preventive Programs in Real Practice. Transitional care interventions (TCIs), defined as care programs to assist patients' seamless transition across care settings (Lee et al. 2022; Rasmussen et al. 2021), have been proven as effective strategies for improving COC, particularly for medically complex patients (Lee et al. 2022; Weeks et al. 2021). TCIs ensure care

continuity after discharge by providing information to patients and care providers for early detection of health issues (Odeh et al. 2019). The widespread TCIs based on the stage of care are interventions at home after patient discharge (Lee et al. 2022). The most common interventions in this category are home visits, telephone follow-ups, and bundled interventions with both home visits and calls, which differ in terms of the follow-up duration, cost, and their effectiveness to prevent the event (Nuckols et al. 2017). Little is known about the types of TCI programs and their effectiveness in Canada (Weeks et al. 2021). Thus, we selected the seven most common TCI programs that are currently being implemented in other countries (mainly the US and UK), for which we had enough information for the purpose of our study. We estimated their costs in Canada following a procedure recommended by a seminal JAMA paper (Nuckols et al. 2017) and applying currency conversion factors to identify the costs in CAD. We also adjusted all costs for inflation based on 2020 CAD using the appropriate Consumer Price Index (Bank of Canada 2021). Table G1 (Appendix G) provides information about the standardization of program costs.

The first intervention program included three structured post-discharge telephone calls by three clinical pharmacists within a 3-month period after hospital discharge, scheduled for within 10 days, one month, and the start of the third month after discharge. The cost of intervention per patient was estimated to be $63 for 90 days of postdischarge follow-up (Odeh et al. 2019), with a standardized cost in 2020 CAD. The next study proposed the three following intervention programs: a predischarge assessment, two post-discharge home visit follow-ups, and two phone calls within 28 days following hospital discharge (Wong, Chow, et al. 2014). Intervention costs for home visits, call visits, and the bundled approach were $196, $89, and $284 per patient, respectively (Wong, So, et al. 2014), with standardized costs in 2020 CAD. The fifth intervention included a home visit shortly after discharge and two or three follow-up telephone calls for six

months after hospital discharge (Gardner et al. 2014). The estimated intervention cost per patient was $465 (Gardner et al. 2014), with a standardized cost in 2020 CAD. The sixth intervention program, which was more comprehensive but also more costly, included assessing patient needs during hospitalization, assessing communication strategies between patients and care providers, reconciling medications, and engaging in home visit follow-up based on a five-step standardized protocol. More information about the intervention can be found in Ornstein et al. (2011). The estimated cost for the program per patient was $614 (Nuckols et al. 2017), with a standardized cost in 2020 CAD. Finally, the last cost-effective postdischarge care transition program, which was designed for a 45-day follow-up after discharge, included three home visits by nurses and several main activities, such as patient-centered health record development, follow-up visit with a physician, coordination of data flow, and patient education (Saleh et al. 2012). The estimated program cost per individual was $1,532 (Saleh et al. 2012), with a standardized cost in 2020 CAD. Nearly all the abovementioned interventions sought to ensure COC after hospitalization with providers.

Odeh et al. (2019) have reported a risk difference of 24.4% after implementing the TCI intervention within 90 days with a cost of $63 (readmission rate of control group: 49.6% vs. readmission rate 25.2% for the intervention group). A recently published systematic review suggested that the intensity of TCIs influences readmission rates because high-intensity interventions generally have a more substantial impact than low-intensity interventions do (Rasmussen et al. 2021). Therefore, we considered 24% as the base effectiveness rate for the least costly intervention, with a cost of $63, and assumed that the effectiveness rate changes linearly with intervention cost (Gupta et al. 2020); this was supported by the results of a meta-analysis suggesting that the effectiveness of a TCI program increases linearly with program costs (with a

rate of 0.7% per $100 investment; Nuckols et al. 2017). Our decision-making framework can now be used to determine the optimal actions for each preventive program to mitigate the fragmentation of care, considering both the intervention cost and its effectiveness in the presence of budget constraints.

Uncapacitated Implementation of Real Practice Intervention Programs. We first considered the case where we had enough budget to implement the preventive intervention for all cases which are predicted as "positive" by ML (i.e., TP+FP). In this analysis, we considered the same budget for the random strategy. Given the available preventive programs, we are interested in knowing whether the ML strategy is more valuable in practice than the random clinical strategy. Considering a sample size of $\mathrm{N} = 100{,}000$ patients and total positive cases predicted by our model ($\mathrm{TP} + \mathrm{FP} \sim 25\mathrm{K}$), Table 3 shows the economic benefits that could be accrued using ML compared with using the same budget to randomly assign each of the above TCIs among patients. As can be seen, the ML strategy exceeded the random strategy (for all TCIs) in terms of cost reductions. Accordingly, it suggested a minimum savings of $3.6 million for the least expensive intervention ($CI^1 = \$63$) and a maximum savings of $5.2 million for the most expensive intervention ($CI^7 = \$1{,}532$) for 100 thousand patients per year (Table 3). Although the cost saving incurred by implementing the expensive programs appears somewhat higher compared with the less expensive ones, this is undesirable because the cost of program implementation substantially increases and might not be affordable for most hospitals ($CI^7: saving = \$5.2\ M\ ;\ cost = \$38\ M$) compared with ($CI^1: saving = \$3.5\ M\ ;\ cost = \$1.5\ M$). Moreover, our results suggest that the ML cost savings increase linearly with the number of patients (Figure G1).

As illustrated in section 3.5, our ML-based system provides fair recommendations either before or after applying the postprocessing correction method. Moreover, our recommendations can be tailored to the patient attributes, such as sensitive attributes (i.e., personalized recommendation).

Table 3. ML cost saving relative to the random strategy for different programs

Programs	CI^j (\$)	e (%)	ML budget (M\$)	Random strategy (M\$)	ML strategy (M\$)	Absolute Cost Savings (M\$)	Relative Cost Saving (%)
Odeh et al.							
90-day Phone call	63.23	24.0	1.58	967.85	964.20	3.65	0.377
Wong et al.							
Phone call	88.45	24.2	2.20	968.45	964.77	3.67	0.379
Home visit	195.53	24.9	4.87	970.98	967.20	3.78	0.390
Combined	283.98	25.5	7.08	973.08	969.20	3.87	0.398
Gardner et al.							
Home visit & Phone call	465.10	26.8	11.59	977.36	973.29	4.07	0.417
Ornstein et al.							
HBPC program	614.41	27.9	15.31	980.88	976.64	4.24	0.432
Saleh et al							
PDCT program	1,532.24	34.3	38.19	1002.59	997.38	5.21	0.520

HBPC: home-based primary care; PDCT: post-discharge care transition

Capacitated Implementation of Real Practice Intervention Programs. We also analyzed optimal decisions under the restricted capacity/budget, which leads to the selective implementation of the ML-based strategy because we can no longer implement the interventions on all cases that are predicted "positive" by ML. Therefore, under capacitated ML implementation, patients must be randomly selected from the predicted positive cases (i.e., TP+FP) to meet the budget constraint. Figure 11 illustrates the cost savings of implementing the ML-based strategy compared with randomly implementing the seven TCIs described above (with different cost and effectiveness rates) when the intervention budget changes from \$50,000 to \$3,000,000. Intuitively, when interventions are not too costly (e.g., CI^1: \$63 and CI^2: \$89), the superiority of the ML-based strategy increases rapidly at first, then after a certain point slowdown again when the budget increases, whereas the ML-based strategy always outperforms the random strategy for more

expensive TCIs (e.g., CI^{3-7} [ranging from CI^3 =\$196 to CI^7 =\$1,532]), and the superiority of the ML-based intervention increases linearly with the budget. The results depend on the program effectiveness rate, which varied between 24% and 34% for the least expensive TCI (CI^1) to the most expensive program (CI^7) in our analysis. It should be recalled that our decision analysis framework consists of three key components, i.e., intervention, fragmented readmission cost, and nonfragmented readmission cost (please see Equations 3-4); therefore, the budget (or capacity) mainly affects the readmission cost terms. The first component is similar for the two strategies until TP+FP because these two strategies require the same budget for the intervention if the number of target patients receiving the intervention is less than the total positive cases (i.e., TP+FP). After this critical point, the total intervention cost increases under the random strategy, but it stays the same for the ML-based strategy. Therefore, any differences are attributed to the difference in the second and third components, i.e., the readmission costs.

We also observed that more investment in the intervention is beneficial when an intervention is cheap enough (e.g., CI^1: \$63 and CI^2: \$89). This increases the likelihood of capturing eligible patients for the intervention at a low cost to prevent fragmented readmission, which costs around \$2,000 more than nonfragmented readmission. However, it is not desirable to engage in a costly intervention (CI^7= \$1532) to randomly find an FN case that ML failed to identify to save around \$2,000. Hence, for quite costly interventions, ML always performs better. Figure 10 shows this trend, where the ML cost saving diminishes for any budget greater than \$1.6 million and \$2.2 million for CI^1 and CI^2, indicating critical points for these two programs given a 100K sample size per year. We also note that the cost savings linearly change with the sample size. The sensitivity results on the sample size (changing the sample size to 1 million patients) are available in Appendix G (Figure G2). The results indicated the same pattern.

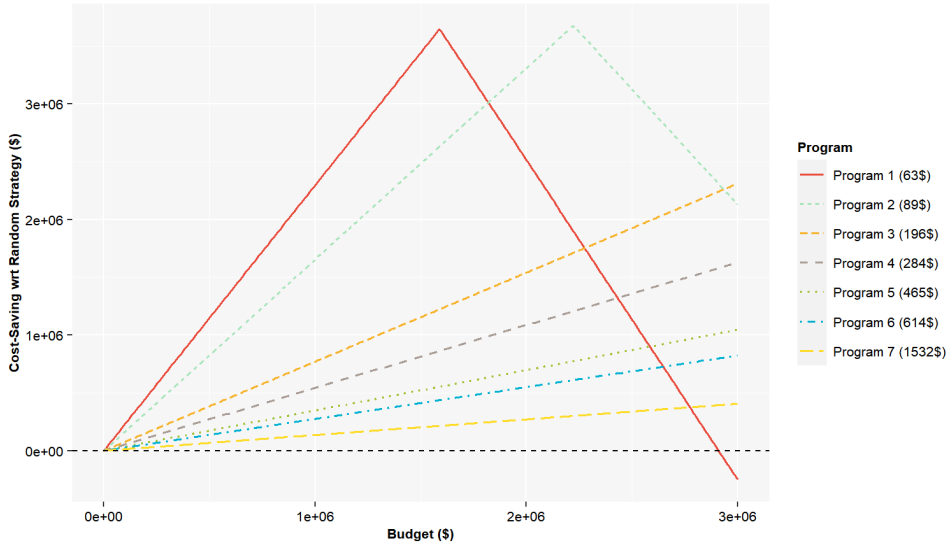


Figure 11. Trend of ML cost savings accrued by increasing the budget for different programs (capacitated analysis)

4.6. Robustness Check on the Cause of COC

Some groups of patients are obliged to go to a different hospital for better/different care, which has led to a natural difference between fragmented and non-fragmented. For example, cancer patients have complex care requirements and specialized expertise, forcing them to visit alternative hospitals, not the hospital where their first treatment is received (Grewal et al. 2019). If this is the case, then the results might be affected by removing such cases. Therefore, to evaluate the robustness of our findings, we removed the records of patients admitted with cancer as the main diagnosis (89,417 observations) and re-conducted the analyses. The results (Appendix H) indicated that there are no significant changes in the cost of readmissions compared to the original cohort ($\bar{C}_{Nonfragmented}$: \$9,319 $vs.$ \$9,448 ; $\bar{C}_{Fragmented}$: 11,263 $vs.$ \$11,344) and fragmented readmission rate (38.3% vs. 38.5%). Hence, the difference between the cost savings —cost of the ML-based strategy implementation compared to the existing clinical strategy—is not tangible (CI^1: $saving$ = \$3.68 $million$ $vs.$ \$3.65; CI^7: $saving$: \$5.26 million $vs.$ \$5.21 $million$).

5. Concluding Remarks and Discussions

Principal Findings. Using extensive data from the older Canadian population, in this retrospective study, we showed the feasibility of predicting fragmented readmissions—that is, unplanned readmission to a different hospital than the index admission—within 30 days after hospital discharge using ML algorithms in the context of the aging population. Using statistical models, we also assessed the possible association between fragmented care and the risk of clinical and cost-related outcomes. This study provided firsthand evidence of accurate fragmented readmission prediction via robust comparisons of multiple predictive models using different performance measures. The predictive models using boosting and bagging algorithms exhibited better performance compared with LR, which is widely used in health service research. The performance measures of the stacked model based on boosting and bagging algorithms as base learners and LR as meta-learner were similar to those of the single learning algorithms. However, this approach was computationally intensive and required more time to train; at the same time, all the constructed models generalized well on the unseen data, indicating a lack of overfitting. We also examined our predictive models' economic and practical effects via a decision and fairness analysis, an approach that is more informative/meaningful in clinical settings. We offered insights regarding the cost savings obtained through ML and suggested a predictive–prescriptive framework that ensures nondiscriminatory decisions regarding the protected groups.

Managerial and Policy Implications. Evidence for accurately predicting the risk of fragmented readmission is essential and has important implications for policy and practice. First, our data-driven risk prediction tool has potential utility in facilitating the screening of patients at risk of fragmented readmission. Patients detected by our predictive model as being at high risk of fragmented readmission could be followed up more closely to mitigate the burden of preventable

costs while improving patient clinical outcomes. In the context of prescriptive analytics, our framework can facilitate decision-making for policymakers in a strategic planning process to determine which cost-effective targeted interventions (to prevent fragmented readmission) have the greatest potential for cost savings in light of budget deficits and affordability. At the service level, the result suggests that an optimal screening strategy could be applied to detect patients at risk of fragmented readmission; this could be tailored to patient attributes, thereby providing personalized recommendations. Below, we discuss how our study provides insights into managerial and policy implications. First, as part of our analysis, we made several observations that may be helpful for policymakers at the system level and organizational leaders at the service level to make an optimal decision regarding resource allocation among older patients at risk of CF. Our results indicated that the effectiveness of the ML-based strategy would depend on the performance of risk models in predicting positive and negative cases, the effectiveness of interventions, the cost of the outcome, the rate of outcomes among subgroups, and the degree of risk aversion. Although the utility of applying the ML strategy is high when the model performance is perfect, even moderately accurate models can outperform random strategies in different scenarios, such as targeting costly interventions to patients in the context of the budget constraint. Second, implementing our predictive–prescriptive framework could yield significant financial savings. For example, given the cost-effective preventive programs for CF in our analysis, employing the ML-based strategy for a cohort of 100K patients could save between $3.6 million to $5.2 million annually compared with the available clinical random strategies in practice, with additional indirect savings available with improved efficiency and effectiveness in care delivery. Utilizing our framework might reduce the cost of unnecessary tests and avoidable fragmented readmission by selecting the right interventions for the right patients. The saved cost stems from

our framework implementation and can be redirected toward expenditures that represent a better investment in patient care, such as acquiring imaging machinery needed to meet the demands of the growing aging population. In Canada, excessive wait times for computed tomography (CT)/magnetic resonance imaging (MRI) diagnostics cost the Canadian economy around $5,000 per patient ($4,136 for CT and $5,853 for MRI; Sutherland et al. 2019). Our key message here is that there is significant potential for cost savings via the reduced fragmentation of care.

Third, our system can facilitate shared decision-making in inpatient care settings by providing a platform to communicate information about probable outcomes (e.g., risk of CF) and discharge planning to patients, their families, and care providers across the care continuum so that they can use it for their informed decision making. Patients and their families can be informed of the risk of CF and decide on more engagement in the process of following up after discharge. A study indicated that up to 70% of patients might have changed their decisions and preferences about their care plans after they were fully informed about their probable outcomes (Robinson and Jagsi 2016). Fourth, the Ontario Ministry of Health and Ontario Ministry of Long-Term Care encourage providers and policymakers to include equity considerations in their decision-making to reduce disparities between underserved populations (Greenwood et al. 2018). Our framework further accounts for such fairness considerations, which can have far-reaching policy implications, by allowing the disparity learned from data to be computed and unfair decisions to be corrected in the shift toward a fair ML-based decision. We focused on impact parity and targeting patients based on the predicted risk of fragmented readmission. This contravenes the current random selection in clinical practice, which does not consider patients' eligibility based on the risk of adverse events (i.e., fragmented readmission). We followed equal opportunity criteria and considered the patient's needs in terms of their service utilization and the event frequency among subgroups. Our

framework can inform resource planning decisions at the systems level and influence actions that affect patient-centered care transition at the service level. Finally, our findings motivate efforts to improve CF and enhance the sharing of clinical records across healthcare institutions, at least through a connection between rural and urban hospitals. Policymakers could seek to address COC issues by providing financial incentives for care providers or healthcare managers together with personalized patient education to encourage patients to refer to the same healthcare institute (when possible), especially for the medically complex aging population, which experiences more care transitions.

Policy Implications for Delayed Discharge Patients. The results of this study have important implications for addressing delayed discharge among older adults. The patients, who are also referred to as ALC patients, receive their needed care with a delay until they are transferred to the most appropriate alternate facility. This delay undermines the "seamless care" component of COC. Moreover, the readmission of ALC patients after their discharge to a different hospital can amplify the impact of CF among this group. The issue of CF becomes even more serious among ALC patients. In addition, our cost estimations suggest that the additional cost of fragmented readmission incurred by the healthcare system of Ontario compared with nonfragmented readmission is significantly greater among ALC patients than it is among non-ALC patients ($5,131 vs. $1,718). Hence, our results shed light on the importance of considering ALC patients in strategies for improving CF among older adults.

Strengths, Limitations, and Future Works. This study presented robust analytics of the fragmented readmission incidence using several ML algorithms and various predictive accuracy measures based on large longitudinal data. However, our study is not without limitations. First, this was a retrospective cohort study with limited control over data collection. Second, we were

unable to use some imbalanced data processing techniques, such as the synthetic minority oversampling technique, because of computational resource challenges. The use of such techniques might enhance the model's predictive performance. Although we acknowledge this weakness, we used two other classic resampling approaches to investigate the impact of imbalanced data on our models' performance. Third, there may have been some other confounders that were not included in our data, such as hospital-level characteristics (e.g., size of hospital/ number of beds, teaching status, hospital location; Brooke et al. 2015; McCrum et al. 2020); therefore, the performance of the predictive models might be affected by adding these risk factors in our models. Further, the distance between the care setting and the patient's home remains a risk factor for non-index readmission but was not included in the data (Juo et al. 2019b). Exploration of the association between COC and this risk factor appears to be an especially worthy endeavor. We acknowledge that although readmission to an index hospital might improve information continuity and avoid duplication in clinical tests, in-hospital therapies, or interventions, the acuteness of patients' health or their needs for more specialized services may force patients to admit to the closest facility or the more appropriate hospital to meet their needs rather than the same hospital as their recent discharge. Although we have focused on informational COC, our framework can be extended to include other aspects of COC, such as managerial and relational COC, and to investigate the extent to which continuity with both care providers and healthcare institutions can be predicted using ML algorithms. This work has also paved the way for the development of frameworks for other important decision-making problems in health care, such as preventive intervention for the rehospitalization or ED transfers of nursing home residents, which are challenging issues in aging research (Dubucs et al. 2019; Lemoyne et al. 2019). A further prospective study is warranted to validate our system in real-world programs and clinical settings.

In summary, as hospitals seek to improve the quality of care they provide and optimize the use of scarce resources, determining ways to optimize *where* care is received is critical. Although health systems are increasingly trying to integrate services and information across sectors and providers, the fact remains that information on patients' care episodes continues to sit largely within siloed institutions. Given the lack of data sharing across providers and sectors, determining ways to improve continuity of care by way of patients utilizing a similar hospital for each admission may help to improve care, understanding of patient needs over time, reduce waste, and create a more seamless experience for patients and their families. Targeting patients at risk of discontinuity of care through the model we have proposed can help to facilitate better use of resources, including preventative options in the community setting.

References

Agha, L., B. Frandsen, J. B. Rebitzer. 2019. Fragmented division of labor and healthcare costs: Evidence from moves across regions. *J. Public Econ.* **169**: 144–159.

Ahuja, V., C. A. Alvarez, B. R. Staats. 2020. Maintaining continuity in service: An empirical examination of primary care physicians. *Manuf. Serv. Oper. Manag.*

Ahuja, V., C. Alvarez, B. R. Staats. 2019. How Continuity in Service Impacts Process Variability: Evidence from a Primary Care Setting. *Available SSRN 3440355*.

Akobeng, A. K. 2005. Understanding randomised controlled trials. *Arch. Dis. Child.* **90**(8): 840–844.

Ando, T., O. Adegbala, P. Villablanca, E. Akintoye, S. Ashraf, M. Shokr, F. Siddiqui, H. Takagi, C. L. Grines, L. Afonso. 2019. Incidence and predictors of readmissions to non-index hospitals after transcatheter aortic valve replacement and the impact on in-hospital outcomes: from the nationwide readmission database. *Int. J. Cardiol.* **292**: 50–55.

Ang, E., S. Kwasnick, M. Bayati, E. L. Plambeck, M. Aratow. 2016. Accurate emergency department wait time prediction. *Manuf. Serv. Oper. Manag.* **18**(1): 141–156.

Anselmi, L., M. Lagarde, K. Hanson. 2015. Equity in the allocation of public sector financial resources in low-and middle-income countries: a systematic literature review. *Health Policy Plan.* **30**(4): 528–545.

Bank of Canada. 2021. Annual average exchange rates. Available at https://www.bankofcanada.ca/rates/related/inflation-calculator/?page_moved=1.

Barker, I., A. Steventon, S. R. Deeny. 2017. Association between continuity of care in general practice and hospital admissions for ambulatory care sensitive conditions: cross sectional study of routinely collected, person level data. *Bmj* **356**: j84.

Barocas, S., M. Hardt, A. Narayanan. 2017. Fairness in machine learning. *Nips Tutor.* **1**: 2.

Bastani, H., M. Bayati. 2020. Online decision making with high-dimensional covariates. *Oper. Res.* **68**(1): 276–294.

Bertsimas, D., A. O'Hair, S. Relyea, J. Silberholz. 2016. An analytics approach to designing combination chemotherapy regimens for cancer. *Manag. Sci.* **62**(5): 1511–1531.

Bertsimas, D., J. Pauphilet, J. Stevens, M. Tandon. 2021. Predicting inpatient flow at a major hospital using interpretable analytics. *Manuf. Serv. Oper. Manag.*

Bielińska, K., A. Chowaniec, R. Doričić, M. Nowak, M. Orzechowski, M. Ramšak, P. Łuków, A. Muzur, Z. Zupanič-Slavec, F. Steger. 2022. Equal access to healthcare in national legislations: how do Croatia, Germany, Poland, and Slovenia counteract discrimination in healthcare? *BMC Health Serv. Res.* **22**(1): 1–11.

Bjarnadottir, M., D. Anderson, L. Zia, K. Rhoads. 2018. Predicting Colorectal Cancer Mortality: Models to Facilitate Patient-Physician Conversations and Inform Operational Decision Making. *Prod. Oper. Manag.* **27**(12): 2162–2183.

Brauer, D. G., N. Wu, M. R. Keller, S. A. Humble, R. C. Fields, C. W. Hammill, W. G. Hawkins, G. A. Colditz, D. E. Sanford. 2021. Care Fragmentation and Mortality in Readmission after Surgery for Hepatopancreatobiliary and Gastric Cancer: A Patient-Level and Hospital-Level Analysis of the Healthcare Cost and Utilization Project Administrative Database. *J. Am. Coll. Surg.* **232**(6): 921–932.

Brooke, B. S. 2020. Transitioning Care at Discharge. *Princ. Pract. Geriatr. Surg.* 485–506.

Brooke, B. S., P. P. Goodney, L. W. Kraiss, D. J. Gottlieb, M. H. Samore, S. R. Finlayson. 2015. Readmission destination and risk of mortality after major surgery: an observational cohort study. *The Lancet* **386**(9996): 884–895.

Burm, S., K. Boese, L. Faden, S. DeLuca, N. Huda, K. Hibbert, M. Goldszmidt. 2019. Recognising the importance of informal communication events in improving collaborative care. *BMJ Qual. Saf.* **28**(4): 289–295.

Canadian Institute for Health Information. (n.d.).

Chatzimparmpas, A., R. M. Martins, K. Kucher, A. Kerren. 2020. Stackgenvis: Alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. *IEEE Trans. Vis. Comput. Graph.* **27**(2): 1547–1557.

Chekroud, A. M., J. Bondar, J. Delgadillo, G. Doherty, A. Wasil, M. Fokkema, Z. Cohen, D. Belgrave, R. DeRubeis, R. Iniesta. 2021. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry* **20**(2): 154–170.

Chicco, D., G. Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**(1): 1–13.

Chicco, D., N. Tötsch, G. Jurman. 2021. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min.* **14**(1): 1–22.

Chouldechova, A., A. Roth. 2018. The frontiers of fairness in machine learning. *ArXiv Prepr. ArXiv181008810*.

Corbett-Davies, S., E. Pierson, A. Feller, S. Goel, A. Huq. 2017. Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*. 797–806.

Covey, C. 2022. Impact of Risk Perception and Risk Tolerance on Investment Portfolio Decisions.

Crilly, J., W. Chaboyer, M. Wallis. 2006. Continuity of care for acutely unwell older adults from nursing homes. *Scand. J. Caring Sci.* **20**(2): 122–134.

Culyer, A. J. 2007. Equity of what in health care? Why the traditional answers don't help policy-and what to do in the future. *Healthc. Pap.* **8**(Special Issue): 12–26.

Cummings, S. R., D. G. Grady, A. J. Huang. 2022. Designing Randomized Blinded Trials. *Des. Clin. Res.*

Deyo, R. A., D. C. Cherkin, M. A. Ciol. 1992. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J. Clin. Epidemiol.* **45**(6): 613–619.

Doupe, P., J. Faghmous, S. Basu. 2019. Machine learning for health services researchers. *Value Health* **22**(7): 808–815.

Dubucs, X., P. de Souto Barreto, C. L. de Mazieres, D. Lauque, O. Azema, S. Charpentier, Y. Rolland. 2019. The temporal trend in the transfer of older adults to the emergency department for traumatic injuries: a retrospective analysis according to their place of residence. *J. Am. Med. Dir. Assoc.* **20**(11): 1462–1466.

Elixhauser, A., C. Steiner, D. R. Harris, R. M. Coffey. 1998. Comorbidity measures for use with administrative data. *Med. Care* **36**(1): 8–27.

Felder, S., T. Mayrhofer. 2022. Basic Tools in Economics: Preferences, Expected Utility, Risk Aversion and Prudence. *Medical Decision Making*. Springer, 23–52.

Ferrer, X., T. van Nuenen, J. M. Such, M. Coté, N. Criado. 2021. Bias and Discrimination in AI: a cross-disciplinary perspective. *IEEE Technol. Soc. Mag.* **40**(2): 72–80.

Fotouhi, S., S. Asadi, M. W. Kattan. 2019. A comprehensive data level analysis for cancer diagnosis on imbalanced data. *J. Biomed. Inform.* **90**: 103089.

Fox, M. T., M. Persaud, I. Maimets, D. Brooks, K. O'Brien, D. Tregunno. 2013. Effectiveness of early discharge planning in acutely ill or injured hospitalized older adults: a systematic review and meta-analysis. *BMC Geriatr.* **13**(1): 70.

Fu, R., M. Aseri, P. Singh, K. Srinivasan. 2021. "Un" fair machine learning algorithms. *Manag. Sci.*

Fu, R., Y. Huang, P. V. Singh. 2020. Artificial intelligence and algorithmic bias: Source, detection, mitigation, and implications. *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*. INFORMS, 39–63.

Gardner, R., Q. Li, R. R. Baier, K. Butterfield, E. A. Coleman, S. Gravenstein. 2014. Is implementation of the care transitions intervention associated with cost avoidance after hospital discharge? *J. Gen. Intern. Med.* **29**(6): 878–884.

Ghassemi, M., E. O. Nsoesie. 2022. In medicine, how do we machine learn anything real? *Patterns* **3**(1): 100392.

Gilbert, T., J. Neuburger, J. Kraindler, E. Keeble, P. Smith, C. Ariti, S. Arora, A. Street, S. Parker, H. C. Roberts. 2018. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *The Lancet* **391**(10132): 1775–1782.

Goldstein, L. B., G. P. Samsa, D. B. Matchar, R. D. Horner. 2004. Charlson Index comorbidity adjustment for ischemic stroke outcome studies. *Stroke* **35**(8): 1941–1945.

Greenwood, M., S. de Leeuw, N. Lindsay. 2018. Challenges in health equity for Indigenous peoples in Canada. *The Lancet* **391**(10131): 1645–1648.

Gulliford, M., S. Naithani, M. Morgan. 2006. What is' continuity of care'? *J. Health Serv. Res. Policy* **11**(4): 248–250.

Gupta, S., R. Sutradhar, Q. Li, U. Athale, M. Bassal, V. Breakey, P. J. Gibson, S. Patel, M. Silva, V. Zabih. 2021. Health care utilisation and costs associated with different treatment protocols for newly diagnosed childhood acute lymphoblastic leukaemia: A population-based study in Ontario, Canada. *Eur. J. Cancer* **151**: 126–135.

Gupta, V., B. R. Han, S.-H. Kim, H. Paek. 2020. Maximizing intervention effectiveness. *Manag. Sci.* **66**(12): 5576–5598.

Haggerty, J. L., R. J. Reid, G. K. Freeman, B. H. Starfield, C. E. Adair, R. McKendry. 2003. Continuity of care: a multidisciplinary review. *Bmj* **327**(7425): 1219–1221.

Hardt, M., E. Price, N. Srebro. 2016. Equality of opportunity in supervised learning. *Adv. Neural Inf. Process. Syst.* **29**.

Hassen, N., A. Lofters, S. Michael, A. Mall, A. D. Pinto, J. Rackal. 2021. Implementing anti-racism interventions in healthcare settings: a scoping review. *Int. J. Environ. Res. Public. Health* **18**(6): 2993.

Hirji, S. A., E. D. Percy, C. K. Zogg, M. Vaduganathan, S. Kiehm, M. Pelletier, T. Kaneko. 2020. Thirty-Day Nonindex Readmissions and Clinical Outcomes After Cardiac Surgery. *Ann. Thorac. Surg.*

Jin, G. Z., A. Lee, S. F. Lu. 2022. Patient Routing to Skilled Nursing Facilities: The Consequences of the Medicare Reimbursement Rule. *Manag. Sci.*

Juo, Y.-Y., Y. Sanaiha, U. Khrucharoen, B. H. Chang, E. Dutson, P. Benharash. 2019a. Care fragmentation is associated with increased short-term mortality during postoperative readmissions: A systematic review and meta-analysis. *Surgery* **165**(3): 501–509.

Juo, Y.-Y., Y. Sanaiha, U. Khrucharoen, B. H. Chang, E. Dutson, P. Benharash. 2019b. Care fragmentation is associated with increased short-term mortality during postoperative readmissions: A systematic review and meta-analysis. *Surgery* **165**(3): 501–509.

Kajaria-Montag, H., M. Freeman, S. Scholtes. 2021. Continuity of Care Increases Clinical Productivity in Primary Care.

Kallus, N., X. Mao, A. Zhou. 2022. Assessing algorithmic fairness with unobserved protected class using data combination. *Manag. Sci.* **68**(3): 1959–1981.

Kaltenborn, Z., K. Paul, J. D. Kirsch, M. Aylward, E. A. Rogers, M. T. Rhodes, M. G. Usher. 2021. Super fragmented: a nationally representative cross-sectional study exploring the fragmentation of inpatient care among super-utilizers. *BMC Health Serv. Res.* **21**(1): 1–10.

Kamalzadeh, H., V. Ahuja, M. Hahsler, M. E. Bowen. 2021. An Analytics-Driven Approach for Optimal Individualized Diabetes Screening. *Prod. Oper. Manag.* **30**(9): 3161–3191.

Keskinocak, P., N. Savva. 2020. A review of the healthcare-management (modeling) literature published in Manufacturing & Service Operations Management. *Manuf. Serv. Oper. Manag.* **22**(1): 59–72.

Kong, L., K. Hu, R. Verma. 2022. Service chains' operational strategies: Standardization or customization? Evidence from the nursing home industry. *Manuf. Serv. Oper. Manag.*

Kothari, A. N., V. M. Loy, S. A. Brownlee, Y. Ezure, C. Schaidle-Blackburn, S. J. Cotler, D. di Sabato, P. C. Kuo, A. D. Lu. 2017. Adverse effect of post-discharge care fragmentation on outcomes after readmissions after liver transplantation. *J. Am. Coll. Surg.* **225**(1): 62–67.

Lawson, C., H. Crothers, S. Remsing, I. Squire, F. Zaccardi, D. Md, L. Bernhardt, K. Reeves, R. Lilford, K. Khunti. 2021. Trends in 30-day readmissions following hospitalisation for heart failure by sex, socioeconomic status and ethnicity. *EClinicalMedicine* **38**: 101008.

Lee, J. Y., Y. S. Yang, E. Cho. 2022. Transitional care from hospital to home for frail older adults: A systematic review and meta-analysis. *Geriatr. Nur. (Lond.)* **43**: 64–76.

Leevy, J. L., T. M. Khoshgoftaar, R. A. Bauder, N. Seliya. 2018. A survey on addressing high-class imbalance in big data. *J. Big Data* **5**(1): 1–30.

Lemoyne, S. E., H. H. Herbots, D. De Blick, R. Remmen, K. G. Monsieurs, P. Van Bogaert. 2019. Appropriateness of transferring nursing home residents to emergency departments: a systematic review. *BMC Geriatr.* **19**(1): 1–9.

Lotz, A. 2019. Use of a Communication Tool to Improve the Perception of Continuity of Care.

Lu, S. F., L. X. Lu. 2017. Do mandatory overtime laws improve quality? Staffing decisions and operational flexibility of nursing homes. *Manag. Sci.* **63**(11): 3566–3585.

Lu, S. F., K. Serfes, G. Wedig, B. Wu. 2021. Does competition improve service quality? The case of nursing homes where public and private payers coexist. *Manag. Sci.*

Maimon, O., L. Rokach. 2005. Data mining and knowledge discovery handbook.

Matheson FI, van Ingen. 2018. 2016 Ontario marginalization index. *Tor. St Michael's Hosp. 2018 Jt. Publ. Public Health Ont.*

McAlister, F. A., E. Youngson, P. Kaul. 2017. Patients with heart failure readmitted to the original hospital have better outcomes than those readmitted elsewhere. *J. Am. Heart Assoc.* **6**(5): e004892.

McCrum, M. L., A. R. Cannon, C. M. Allen, A. P. Presson, L. C. Huang, B. S. Brooke. 2020. Contributors to increased mortality associated with care fragmentation after emergency general surgery. *JAMA Surg.* **155**(9): 841–848.

Meijboom, B. R., S. J. Bakx, G. P. Westert. 2010. Continuity in health care: lessons from supply chain management. *Int. J. Health Plann. Manage.* **25**(4): 304–317.

Merdan, S., C. Barnett, B. T. Denton, J. E. Montie, D. Miller. 2017. *Data analytics for optimal detection of metastatic prostate cancer*. Technical report, University of Michigan.

Mišić, V. V., G. Perakis. 2020. Data analytics in operations management: A review. *Manuf. Serv. Oper. Manag.* **22**(1): 158–169.

MIT News. 2021, March 23. 3 Questions: Artificial intelligence for health care equity. *MIT News Mass. Inst. Technol.*

Mohliver, A., A. Ody-Brasier. 2022. Religious Affiliation and Wrongdoing: Evidence from US Nursing Homes. *Manag. Sci.*

Morse, K. E., S. C. Bagley, N. H. Shah. 2020. Estimate the hidden deployment cost of predictive models to improve patient care. *Nat. Med.* **26**(1): 18–19.

Nuckols, T. K., E. Keeler, S. Morton, L. Anderson, B. J. Doyle, J. Pevnick, M. Booth, R. Shanman, A. Arifkhanova, P. Shekelle. 2017. Economic evaluation of quality improvement interventions designed to prevent hospital readmission: a systematic review and meta-analysis. *JAMA Intern. Med.* **177**(7): 975–985.

Odeh, M., C. Scullin, G. Fleming, M. G. Scott, R. Horne, J. C. McElnay. 2019. Ensuring continuity of patient care across the healthcare interface: Telephone follow-up post-hospitalization. *Br. J. Clin. Pharmacol.* **85**(3): 616–625.

Ornstein, K., K. L. Smith, D. H. Foer, M. T. Lopez-Cantor, T. Soriano. 2011. To the hospital and back home again: a nurse practitioner-based transitional care program for hospitalized homebound people. *J. Am. Geriatr. Soc.* **59**(3): 544–551.

Passman, J., R. Xiong, J. Hatchimonji, E. Kaufman, C. Sharoky, W. Yang, B. P. Smith, D. Holena. 2020. Readmissions after injury: is fragmentation of care associated with mortality? *J. Surg. Res.* **250**: 209–215.

Poly, T. N., M. M. Islam, M. S. Muhtar, H.-C. Yang, P. A. A. Nguyen, Y.-C. J. Li. 2020. Machine learning approach to reduce alert fatigue using a disease medication–related clinical decision support system: model development and validation. *JMIR Med. Inform.* **8**(11): e19489.

Qi, M., O. Cahan, M. A. Foreman, D. M. Gruen, A. K. Das, K. P. Bennett. 2021. Quantifying representativeness in randomized clinical trials using machine learning fairness metrics. *JAMIA Open* **4**(3): ooab077.

Quan, H., B. Li, C. M. Couris, K. Fushimi, P. Graham, P. Hider, J.-M. Januel, V. Sundararajan. 2011. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am. J. Epidemiol.* **173**(6): 676–682.

Queenan, C., K. Cameron, A. Snell, J. Smalley, N. Joglekar. 2019. Patient heal thyself: reducing hospital readmissions with technology-enabled continuity of care and patient activation. *Prod. Oper. Manag.* **28**(11): 2841–2853.

Radinmanesh, M., F. Ebadifard Azar, B. Najafi, R. Majdzadeh. 2021. A review of appropriate indicators for need-based financial resource allocation in health systems. *BMC Health Serv. Res.* **21**(1): 1–12.

Rajkomar, A., M. Hardt, M. D. Howell, G. Corrado, M. H. Chin. 2018. Ensuring fairness in machine learning to advance health equity. *Ann. Intern. Med.* **169**(12): 866–872.

Rasmussen, L. F., L. B. Grode, J. Lange, I. Barat, M. Gregersen. 2021. Impact of transitional care interventions on hospital readmissions in older medical patients: A systematic review. *BMJ Open* **11**(1): e040057.

Richardson, A. 2022. Biased data lead to biased algorithms. *Can. Med. Assoc. J.* **194**(9): E341.

Robinson, J. D., R. Jagsi. 2016. Physician-patient communication—an actionable target for reducing overly aggressive care near the end of life. *JAMA Oncol.* **2**(11): 1407–1408.

Rosenberg, N. A., D. M. Zulman. 2020. Measures of care fragmentation: Mathematical insights from population genetics. *Health Serv. Res.* **55**(2): 318–327.

Saleh, S. S., C. Freire, G. Morris-Dickinson, T. Shannon. 2012. An effectiveness and cost-benefit analysis of a hospital-based discharge transition program for elderly Medicare recipients. *J. Am. Geriatr. Soc.* **60**(6): 1051–1056.

Samorani, M., S. L. Harris, L. G. Blount, H. Lu, M. A. Santoro. 2021. Overbooked and overlooked: machine learning and racial bias in medical appointment scheduling. *Manuf. Serv. Oper. Manag.*

Saultz, J. W. 2003. Defining and measuring interpersonal continuity of care. *Ann. Fam. Med.* **1**(3): 134–143.

Senot, C. 2019. Continuity of care and risk of readmission: An investigation into the healthcare journey of heart failure patients. *Prod. Oper. Manag.* **28**(8): 2008–2030.

Seyyed-Kalantari, L., H. Zhang, M. McDermott, I. Y. Chen, M. Ghassemi. 2021. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat. Med.* **27**(12): 2176–2182.

Sharma, N., R. Schwendimann, O. Endrich, D. Ausserhofer, M. Simon. 2021. Comparing Charlson and Elixhauser comorbidity indices with different weightings to predict in-hospital mortality: an analysis of national inpatient data. *BMC Health Serv. Res.* **21**(1): 13.

Snow, K., K. Galaviz, S. Turbow. 2020. Patient outcomes following interhospital care fragmentation: a systematic review. *J. Gen. Intern. Med.* **35**(5): 1550–1558.

Suresh, K. P. 2011. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. *J. Hum. Reprod. Sci.* **4**(1): 8.

Sutherland, G., N. Russell, R. Gibbard, A. Dobrescu. 2019. *The Value of Radiology, Part II.* (Ottawa: The Conference Board of Canada, 2019.).

Teo, K., C. W. Yong, F. Muhamad, H. Mohafez, K. Hasikin, K. Xia, P. Qian, S. Dhanalakshmi, N. Priya Utama, K. W. Lai. 2021. The promise for reducing healthcare cost with predictive model: An analysis with quantized evaluation metric on readmission. *J. Healthc. Eng.* **2021**.

Tran, Dat T., R. S. Sheldon, P. Kaul, R. K. Sandhu. 2020. The current and future hospitalization cost burden of syncope in Canada. *CJC Open* **2**(4): 222–228.

Tran, Diem TT, J. V. Tu, J.-Y. Dupuis, A. Bader Eddeen, L. Y. Sun. 2018. Association of frailty and long-term survival in patients undergoing coronary artery bypass grafting. *J. Am. Heart Assoc.* **7**(15): e009882.

Van Calster, B., L. Wynants, J. F. Verbeek, J. Y. Verbakel, E. Christodoulou, A. J. Vickers, M. J. Roobol, E. W. Steyerberg. 2018. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur. Urol.* **74**(6): 796–804.

Van Walraven, C., N. Oake, A. Jennings, A. J. Forster. 2010. The association between continuity of care and outcomes: a systematic and critical review. *J. Eval. Clin. Pract.* **16**(5): 947–956.

Verma, S., J. Rubin. 2018. Fairness definitions explained. *2018 ieee/acm international workshop on software fairness (fairware)*. IEEE, 1–7.

Vickers, A. J., E. B. Elkin. 2006. Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making* **26**(6): 565–574.

Vickers, A. J., B. van Calster, E. W. Steyerberg. 2019. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn. Progn. Res.* **3**(1): 1–8.

Wang, S.-L., J.-D. Chen, P. A. Stirpe, T.-P. Hong. 2011. Risk-neutral evaluation of information security investment on data centers. *J. Intell. Inf. Syst.* **36**(3): 329–345.

Weeks, L. E., B. Barber, E. S. MacDougall, M. Macdonald, R. Martin-Misener, G. Warner. 2021. An exploration of Canadian transitional care programs for older adults. *Healthcare Management Forum* (Vol. 34). SAGE Publications Sage CA: Los Angeles, CA, 163–168.

Weiner, M. G., W. Sheikh, H. P. Lehmann. 2018. Interactive cost-benefit analysis: providing real-world financial context to predictive analytics. *AMIA Annual Symposium Proceedings* (Vol. 2018). American Medical Informatics Association, 1076.

Wiens, J., S. Saria, M. Sendak, M. Ghassemi, V. X. Liu, F. Doshi-Velez, K. Jung, K. Heller, D. Kale, M. Saeed. 2019. Do no harm: a roadmap for responsible machine learning for health care. *Nat. Med.* **25**(9): 1337–1340.

Wodchis, W. P., K. Bushmeneva, M. Nikitovic, I. McKillop. 2013. Guidelines on person-level costing using administrative databases in Ontario.

Wong, F. K. Y., S. K. Y. Chow, T. M. F. Chan, S. K. F. Tam. 2014. Comparison of effects between home visits with telephone calls and telephone calls only for transitional discharge support: a randomised controlled trial. *Age Ageing* **43**(1): 91–97.

Wong, F. K. Y., C. So, J. Chau, A. K. P. Law, S. K. F. Tam, S. McGhee. 2014. Economic evaluation of the differential benefits of home visits with telephone calls and telephone calls only in transitional discharge support. *Age Ageing* **44**(1): 143–147.

Xu, H., N. Zhang. 2021. Implications of data anonymization on the statistical evidence of disparity. *Manag. Sci.*

Zychlinski, N., A. Mandelbaum, P. Momčilović, I. Cohen. 2020. Bed blocking in hospitals due to scarce capacity in geriatric institutions—cost minimization via fluid models. *Manuf. Serv. Oper. Manag.* **22**(2): 396–411.

Supplementary Appendix

Appendix A. Prognostication of the Fragmented Readmission (Methods, Results, Discussion)

Methods. To compare the difference in outcomes between fragmented and nonfragmented readmissions, we performed multivariable LR to determine the outcome difference between the two types of readmission. We conducted regression models for clinical outcomes, including in-hospital mortality and resource utilization outcomes (e.g., LOS and RIW following readmission), adjusted for all covariates, with a separate multivariable model for each outcome. The outcomes were defined as follows: (i) mortality vs. survival, (ii) LOS more than two weeks vs. less than two weeks, and (iii) RIW more than one vs. less than one within 30 days after index admission. We used the odds ratio (OR) to measure association relative to "outcome = 0" as the reference outcome (i.e., survival, LOS less than two weeks, or RIW less than 1 for three outcomes). Sensitivity analyses were performed to ascertain whether the results changed across different outcome window lengths, including 60, 90, and 360 days. We also examined the risk of outcomes separately for each fiscal year, ranging from 2004 to 2016.

Results. We assessed the prognostic value of the fragmented readmission in predicting the clinical and resource utilization outcomes following readmission. On multivariate analysis, fragmented readmission was strongly associated with the increase in both resource utilization outcomes (OR^{30d}_{LOS} = 1.12, 95% Confidence Interval (CI) = 1.11–1.13, p-value <0.001; OR^{30d}_{RIW} = 1.42, 95% CI = 1.40–1.44, $p < 0.001$), but not with increased in-hospital mortality following readmission ($OR^{30d}_{mortality}$ = 0.98, 95% CI = 0.96–1.00, $p = 0.007$). To evaluate the robustness of our findings, we changed the primary window length (30 days) to 60, 90, and 360 days to explore whether the results would change across different window lengths (Figure A1). Accordingly, it was found that the results were not sensitive to the window length; the risk of undesired outcomes slightly

decreased from window lengths of 30 days to 360 days. Further examination of outcomes' risk for each fiscal year revealed that mortality rates were not significantly different between the two groups of fragmented versus nonfragmented readmissions, except for one fiscal year (fiscal year = 2012). However, the risk of resource utilization outcomes—namely, RIW and LOS—remained significant, with an increasing pattern from 2004 to 2016. Details are available in Figure A2 and Table A1.

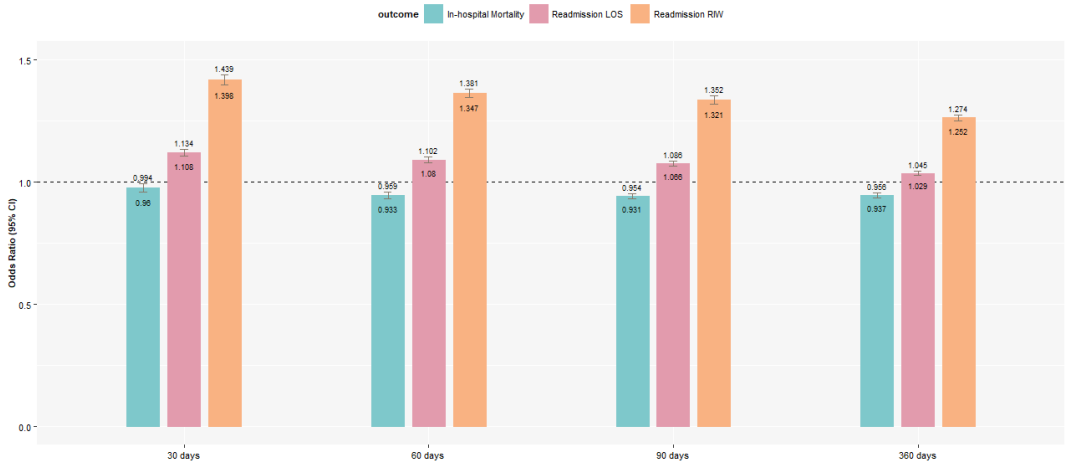


Figure A1. Association of fragmented readmission with outcomes across different window lengths- OR (95% CI)

Table A1. Adjusted and unadjusted Odds Ratios of fragmented readmission within 30 days of index discharge

Outcome	Odds Ratio (95% Confidence Interval) In-hospital Mortality	p-value	Readmission LOS	p-value	Readmission RIW	p-value
Fragmented Readmission:						
Unadjusted Model						
30-days	0.983 (0.968-0.998)	0.030	1.134 (1.123-1.146)	<0.001	1.412 (1.393-1.431)	<0.001
60-days	0.923 (0.912-0.935)	<0.001	1.100 (1.091-1.111)	<0.001	1.347 (1.331-1.362)	<0.001
90-days	0.915 (0.905-0.926)	<0.001	1.081 (1.071-1.090)	<0.001	1.314 (1.300-1.328)	<0.001
360-days	0.898 (0.889-0.909)	<0.001	1.023 (1.016-1.030)	<0.001	1.235 (1.225-1.245)	<0.001
Adjusted Model†						
30-days	0.977 (0.960-0.994)	0.007	1.121 (1.108-1.134)	<0.001	1.418 (1.398-1.439)	<0.001
60-days	0.947 (0.933-0.959)	<0.001	1.091 (1.080-1.102)	<0.001	1.364 (1.347-1.381)	<0.001
90-days	0.942 (0.931-0.954)	<0.001	1.076 (1.066-1.086)	<0.001	1.337 (1.321-1.352)	<0.001
360-days	0.947 (0.937-0.956)	<0.001	1.036 (1.029-1.045)	<0.001	1.263 (1.252-1.274)	<0.001

Note: LOS=length of stay, RIW=resource intensity weight.

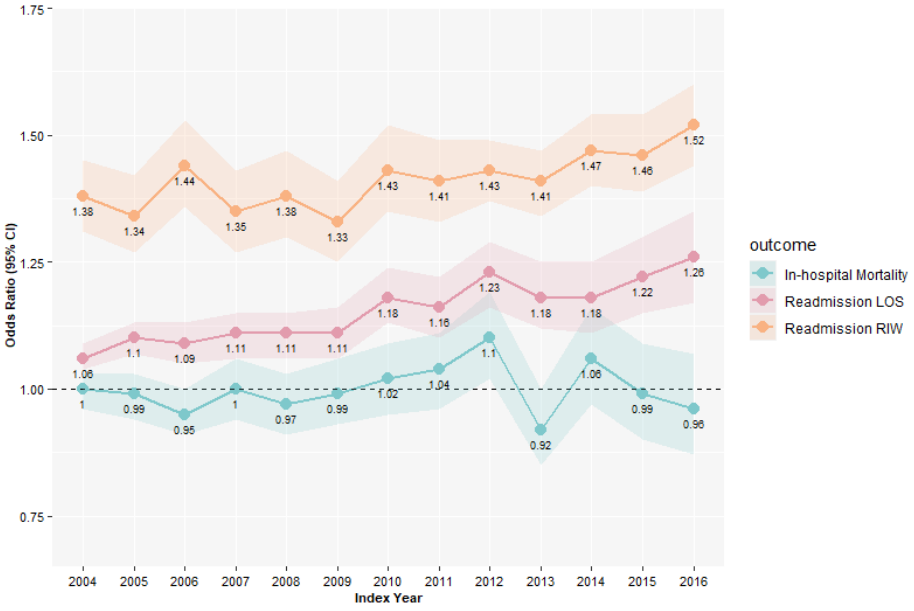


Figure A2. The odds ratio (95% CI) of outcomes in the different index year

Appendix B. Descriptive Analysis

Table B1. Descriptive details regarding the patient characteristics

Characteristics	Fragmented Readmission (n=368,057)	Non-fragmented Readmission (n=585,761)	p-value§	Total (n=953,818)	OR (95% CI)	p-value‡
Age, mean (SD)	74.5 (7.4)	75.5 (7.8)	-	75.1 (7.6)	-	-
65-79	270,996 (74%)	400,884 (68%)	<0.001	671,880 (70%)	Ref.	
>=80	97,061 (26%)	184,877 (32%)	<0.001	281,938 (30%)	0.77 (0.76-0.78)	<0.001
Sex						
Male	178,541 (49%)	284,472 (49%)	0.59	463,013 (49%)	Ref.	
Female	189,516 (51%)	301,289 (51%)		490,805 (51%)	1.00 (0.99-1.01)	0.60
Residency						
Urban	263,922 (72%)	492,966 (84%)	<0.001	756,888 (79%)	Ref.	
Rural	104,135 (28%)	92,795 (16%)		196,930 (21%)	2.10 (2.08-2.12)	<0.001
Material deprivation						
Quintile 1(least)	60,200 (16%)	95,849 (16%)	0.93	156,049 (16%)	Ref.	
Quintile 2	65,984 (18%)	105,982 (18%)	0.04	171,966 (18%)	0.99 (0.98-1.01)	0.22
Quintile 3	73,417 (20%)	116,223 (20%)	0.21	189,640 (20%)	1.01 (0.99-1.02)	0.41
Quintile 4	78,147 (21%)	124,324 (21%)	0.93	202,471 (21%)	1.00 (0.99-1.01)	0.91
Quintile 5(most)	90,309 (25%)	143,383 (25%)	0.52	233,692 (25%)	1.00 (0.99-1.02)	0.52
Residential instability						
Quintile 1(least)	43,522 (12%)	68,860 (12%)	0.31	112,382 (12%)	Ref.	
Quintile 2	60,021 (16%)	95,827 (16%)	0.51	155,848 (16%)	0.99 (0.98-1.01)	0.26
Quintile 3	72,660 (20%)	115,590 (20%)	0.92	188,250 (20%)	0.99 (0.98-1.01)	0.48
Quintile 4	80,277 (22%)	127,783 (22%)	0.97	208,060 (22%)	0.99 (0.98-1.01)	0.43
Quintile 5(most)	111,577 (30%)	177,701 (30%)	0.82	289,278 (30%)	0.99 (0.98-1.01)	0.45
Ethnic concentration						
Quintile 1(least)	95,774 (26%)	152,655 (26%)	0.67	248,429 (26%)	Ref.	
Quintile 2	78,772 (21%)	125,733 (21%)	0.47	204,505 (21%)	1.00 (0.99-1.01)	0.82

Quintile 3	68,588 (19%)	107,594 (19%)	0.00	176,182 (19%)	1.02 (1.00-1.03)	0.01
Quintile 4	61,871 (17%)	99,343 (17%)	0.06	161,214 (17%)	0.99 (0.98-1.01)	0.26
Quintile 5(most)	63,052 (17%)	100,436 (17%)	0.85	163,488 (17%)	1.00 (0.99-1.01)	0.92
Dependency						
Quintile 1(least)	40,374 (11%)	63,678 (11%)	0.13	104,052 (11%)	Ref.	
Quintile 2	52,142 (14%)	83,628 (14%)	0.14	135,770 (14%)	0.98 (0.97-1.00)	0.05
Quintile 3	62,672 (17%)	99,866 (17%)	0.79	162,538 (17%)	0.99 (0.97-1.01)	0.21
Quintile 4	77,165 (21%)	122,855 (21%)	0.93	200,020 (21%)	0.99 (0.98-1.01)	0.23
Quintile 5(most)	135,704 (37%)	215,734 (37%)	0.69	351,438 (37%)	0.99 (0.98-1.01)	0.34
Method of entry						
ED	243,329 (66%)	480,325 (82%)		723,654 (76%)	Ref.	
Non-ED	124,728 (34%)	105,436 (18%)	<0.001	230,164 (24%)	2.34 (2.31-2.36)	<0.001
Acute LOS						
<14 days	316,061 (86%)	509,599 (87%)		825,660 (87%)	Ref.	
≥14 days	51,996 (14%)	76,162 (13%)	<0.001	128,158 (13%)	1.10 (1.09-1.11)	<0.001
ALC LOS						
<14 days	358,789 (97%)	571,883 (98%)		930,672 (98%)	Ref.	
≥14 days	9,268 (3%)	13,878 (2%)	<0.001	23,146 (2%)	1.10 (1.09-1.11)	<0.001
ALC status						
No	358,789 (97%)	571,883 (98%)		930,672 (98%)	Ref.	
Yes	9,268 (3%)	13,878 (2%)	<0.001	23,146 (2%)	1.06 (1.04-1.09)	<0.001
Discharge destination						
Routine (home)	103,140 (28%)	284,500 (49%)		387,640 (41%)	Ref.	
Others	264,917 (72%)	301,261 (51%)	<0.001	566,178 (59%)	2.42 (2.40-2.44)	<0.001
# Comorbidity, mean (SD)	4.58 (3.8)	4.46 (3.6)		4.50 (3.7)	1.009 (1.008-1.01)	<0.001
HFRS, mean (SD)	2.55 (3.4)	2.59 (3.5)		2.59 (3.5)	0.995 (0.994-0.996)	<0.001
# Visited SCU						
0	264,481 (72%)	504,308 (87%)	<0.001	768,789 (81%)	Ref.	
1	91,313 (25%)	72,387 (12%)	<0.001	163,700 (17%)	2.41 (2.37,2.43)	<0.001
≥2	12,263 (3%)	9,066 (2%)	<0.001	21,329 (2%)	2.58 (2.50-2.65)	<0.001
SCU type						
0	264,481 (72%)	504,308 (86%)	<0.001	768,789 (81%)	Ref.	
CICU	21,149 (6%)	13,363 (2%)	<0.001	34,512 (4%)	3.02 (2.95-3.09)	<0.001
MICU	8,909 (2%)	8,386 (1%)	<0.001	17,295 (2%)	2.03 (1.97-2.09)	<0.001
SDSU	3,076 (1%)	2,911 (0.5%)	<0.001	5,987 (1%)	2.01 (1.91-2.12)	<0.001
SICU	2,617 (1%)	2,337 (0.5%)	<0.001	4,954 (1%)	2.14 (2.02-2.26)	<0.001
SDMU	6,593 (2%)	7,736 (1.5%)	<0.001	14,329 (2%)	1.63 (1.57-1.68)	<0.001
CMSICU	48,846 (13%)	38,813 (7%)	<0.001	87,659 (9%)	2.40 (2.37-2.43)	<0.001
Others	12,386 (3%)	7,907 (1.5%)	<0.001	20,293 (2%)	2.99 (2.90-3.07)	<0.001
SCU LOS						
0 h	264,749 (72%)	504,366 (86%)	<0.001	769,115 (81%)	Ref.	
≤72 h	57,001 (15%)	42,062 (7%)	<0.001	99,063 (10%)	2.58 (2.55-2.62)	<0.001
>72h	46,307 (13%)	39,333 (7%)	<0.001	85,640 (9%)	2.24 (2.21-2.28)	<0.001
Top CMG diagnosis:						
1-Acute MI						
No	331,166 (90%)	566,714 (97%)	<0.001	897,880 (94%)	Ref.	
Yes	36,891 (10%)	19,047 (3%)	<0.001	55,938 (6%)	3.31 (3.26-3.37)	<0.001
2-HF						
No	351,449 (95%)	539,195 (92%)	<0.001	890,644 (93%)	Ref.	
Yes	16,608 (5%)	46,566 (8%)	<0.001	63,174 (7%)	0.55 (0.54-0.56)	<0.001
3-Pneumonia						
No	361,787 (98%)	567,679 (97%)	<0.001	929,466 (97%)	Ref	
Yes	6,270 (2%)	18,082 (3%)	<0.001	24,352 (3%)	0.54 (0.53-0.56)	<0.001
4-COPD						
No	357,291 (97%)	538,831 (92%)	<0.001	896,122 (94%)	Ref.	
Yes	10,766 (3%)	46,930 (8%)	<0.001	57,696 (6%)	0.35 (0.34-0.35)	<0.001

5-Gonarthrosis						
No	364,442 (99%)	580,748 (99%)	<0.001	945,190 (99%)	Ref.	
Yes	3,615 (1%)	5,013 (1%)	<0.001	8,628 (1%)	1.15 (1.10-1.20)	<0.001
6-Urinary Disorder						
No	363,765 (99%)	569,988 (97%)	<0.001	933,753 (98%)	Ref.	
Yes	4,292 (1%)	15,773 (3%)	<0.001	20,065 (2%)	0.43 (0.41-0.44)	<0.001
Sepsis						
No	357,215 (97%)	570,828 (97.5%)	<0.001	928,043 (97%)	Ref.	
Yes	10,842 (3%)	14,933 (2.5%)	<0.001	25,775 (3%)	1.16 (1.13-1.19)	<0.001
Top Main patient service:						
1-General Medicine						
No	194,346 (53%)	276,901 (47%)	<0.001	471,247 (49%)	Ref.	
Yes	173,711 (47%)	308,860 (53%)	<0.001	482,571 (51%)	0.43 (0.41-0.44)	<0.001
2-Cardiology						
No	313,553 (85%)	530,854 (91%)	<0.001	844,407 (89%)	Ref.	
Yes	54,504 (15%)	54,907 (9%)	<0.001	109,411 (11%)	1.68 (1.66-1.70)	<0.001
3-Respirology						
No	358,284 (97%)	557,098 (95%)	<0.001	915,382 (96%)	Ref.	
Yes	9,773 (3%)	28,663 (5%)	<0.001	38,436 (4%)	0.53 (0.52-0.54)	<0.001
4-General Surgery						
No	347,094 (94%)	542,993 (93%)	<0.001	890,087 (93%)	Ref.	
Yes	20,963 (6%)	42,768 (7%)	<0.001	63,731 (7%)	0.77 (0.75-0.78)	<0.001
5-Orthopae Surgery						
No	345,498 (94%)	564,467 (96%)	<0.001	909,965 (95%)	Ref.	
Yes	22,559 (6%)	21,294 (4%)	<0.001	43,853 (5%)	1.73 (1.70-1.76)	<0.001
6-Urology						
No	359,855 (98%)	565,010 (96.5%)	<0.001	924,865 (97%)	Ref.	
Yes	8,202 (2%)	20,751 (3.5%)	<0.001	28,953 (3%)	0.62 (0.60-0.64)	<0.001
Surgical Services Flag						
No	299,953 (81.5%)	505,271 (86%)	<0.001	805,224 (84%)	Ref.	
Yes	68,104 (18.5%)	80,490 (14%)	<0.001	148,594 (16%)	1.43 (1.41-1.44)	<0.001
Top CACS Interventions:						
1-PTCA						
No	367,395 (99.8%)	584,852 (99.8%)	<0.001	952,247 (99.8%)	Ref.	
Yes	662 (0.2%)	909 (0.2%)	<0.001	1,571 (0.2%)	1.16 (1.05-1.28)	<0.001
2-Joint Replacement						
No	367,134 (99.7%)	583,942 (99.7%)	<0.001	951,076 (99.7%)	Ref.	
Yes	923 (0.3%)	1,819 (0.3%)	<0.001	2,742 (0.3%)	0.81 (0.75-0.87)	<0.001
3-**Prostate Resection**						
No	367,604 (99.9%)	584,367 (99.8%)	<0.001	951,971 (99.8%)	Ref.	
Yes	453 (0.1%)	1,394 (0.2%)	<0.001	1,847 (0.2%)	0.52 (0.46-0.57)	<0.001
4- **Urinary Tract**						
No	367,768 (99.9%)	585,029 (99.9%)	<0.001	952,797 (99.9%)	Ref.	
Yes	289 (0.1%)	732 (0.1%)	<0.001	1,021 (0.1%)	0.63 (0.55-0.72)	<0.001
5-**Not-Generally Ambulatory**						
No	367,357 (99.8%)	584,553 (99.8%)	0.09	951,910 (99.8%)	Ref.	
Yes	700 (0.2%)	1,208 (0.2%)	0.09	1,908 (0.2%)	0.92 (0.84-1.01)	0.09
Feeding Tube						
No	365,042 (99%)	582,403 (99.4%)	<0.001	947,445 (99%)	Ref.	
Yes	3,015 (1%)	3,358 (0.6%)	<0.001	6,373 (1%)	1.43 (1.36-1.50)	<0.001
Parenteral Nutrition						
No	364,748 (99%)	581,839 (99.3%)	<0.001	946,587 (99.2%)	Ref.	
Yes	3,309 (1%)	3,922 (0.7%)	<0.001	7,231(0.8%)	1.32 (1.28-1.41)	<0.001
Chemotherapy						
No	366,011 (99%)	581,702 (99.3%)	<0.001	947,713 (99%)	Ref.	
Yes	2,046 (1%)	4,059 (0.7%)	<0.001	6,105 (1%)	0.80 (0.76-0.84)	<0.001

Radiotherapy						
No	365,868 (99.4%)	583,000 (99.5%)	<0.001	948,868 (99.5%)	Ref.	
Yes	2,189 (0.6%)	2,761 (0.5%)	<0.001	4,950 (0.5%)	1.26 (1.19-1.34)	<0.001
Vascular Access Device						
No	350,142 (95%)	565,738 (96.6%)	<0.001	915,880 (96%)	Ref.	
Yes	17,915 (5%)	20,023 (3.4%)	<0.001	37,938 (4%)	1.45 (1.42-1.48)	<0.001
Dialysis						
No	361,217 (98%)	574,065 (98%)	<0.001	935,282 (98%)	Ref.	
Yes	6,840 (2%)	11,696 (2%)	<0.001	18,536 (2%)	0.93 (0.90-0.96)	<0.001
Paracentesis						
No	365,667 (99%)	580,198 (99.1%)	<0.001	945,865 (99%)	Ref.	
Yes	2,390 (1%)	5,563 (0.9%)	<0.001	7,953 (1%)	0.68 (0.65-0.72)	<0.001
Pleurocentesis						
No	361,846 (98.3%)	577,016 (98.5%)	<0.001	938,862 (98.4%)	Ref.	
Yes	6,211 (1.7%)	8,745 (1.5%)	<0.001	14,956 (1.6%)	1.13 (1.10-1.17)	<0.001
Tracheostomy						
No	366,210 (99%)	584,936 (99. 9%)	<0.001	951,146 (99.7%)	Ref.	
Yes	1,847 (1%)	825 (0.1%)	<0.001	2,672 (1.3%)	3.58 (3.29-3.88)	<0.001
MV (short)						
No	353,170 (96%)	577,715 (98.6%)	<0.001	930,885	Ref.	
Yes	14,887 (4%)	8,046 (1.4%)	<0.001	22,933	3.03 (2.94-3.11)	<0.001
MV (long)						
No	362,743 (99%)	583,168 (99.6%)	<0.001	945,911 (99.2%)	Ref.	
Yes	5,314 (1%)	2,593 (0.4%)	<0.001	7,907 (0.8%)	3.29 (2.94-3.11)	<0.001
Heart Resuscitation						
No	366,989 (99.7%)	585,288 (99.9%)	<0.001	952,277 (99.8%)	Ref.	
Yes	1,068 (0.3%)	473 (0.1%)	<0.001	1,541 (0.2%)	3.60 (3.14-3.45)	<0.001
Non-invasive Biopsy						
No	360,480 (98%)	570,833 (97.5%)	<0.001	931,313 (97.6%)	Ref.	
Yes	7,577 (2%)	14,928 (2.5%)	<0.001	22,505 (2.4%)	0.80 (0.78-0.83)	<0.001
Per-Orifice Endoscopy						
No	361,075 (98%)	570,092 (97%)	<0.001	931,167 (97.6%)	Ref.	
Yes	6,982 (2%)	15,669 (3%)	<0.001	22,651 (2.4%)	0.70 (0.68-0.72)	<0.001

§ p-value for equality of proportions in two groups of readmissions (fragmented vs. non-fragmented)

‡ p-value of odds ratio (univariate analysis)- outcome: fragmented readmission.

Note: LOS: length of stay, ALC: alternate level of care, HFRS: hospital frailty risk score, SCU: special care unit, ICU: intensive care unit, CICU: coronary ICU, MICU: medical ICU, SDSU: step-down surgical unit, SICU: surgical ICU, SDMU: step-down medical unit, CMSICU: combined medical/surgical ICU, MI: myocardial infarction, HF: heart failure, COPD: chronic obstructive pulmonary disease, CACS: comprehensive ambulatory classification system, PTCA: percutaneous transluminal coronary vessel, MV: Mechanical Ventilation

Appendix C. Predictive Analytics

C.1. Results of five-fold cross-validated predictive performance of ML models

Table C1. Five-fold cross-validated predictive performance of ML algorithms with resampling approaches

Algorithms	Mean (95 % Confidence Interval)					
	AUC	Specificity	Sensitivity/ Recall	Precision	F1	Accuracy
Resampling: No						
NB	0.676 (0.675-0.677)	0.487 (0.479-0.494)	0.773 (0.772-0.774)	0.705 (0.702-0.708)	0.738 (0.736-0.739)	0.663 (0.660-0.665)
LR	0.734 (0.732-0.736)	0.477 (0.475-0.479)	0.857 (0.854-0.861)	0.723 (0.722-0.724)	0.784 (0.783-0.786)	0.710 (0.709-0.712)
CART	0.730 (0.726-0.733)	0.480 (0.472-0.488)	0.870 (0.864-0.875)	0.727 (0.725-0.728)	0.792 (0.790-0.793)	0.719 (0.719-0.719)
XGB	0.745 (0.744-0.746)	0.477 (0.470-484)	0.874 (0.871-0.878)	0.727 (0.724-0.729)	0.794 (0.793-0.794)	0.721 (720-0.722)
RF	**0.750** (0.748-0.752)	0.483 (0.478-0.488)	**0.876** (0.874-0.878)	0.729 (0.727-0.732)	**0.796** (0.794-0.798)	**0.724** (0.722-0.727)
ADA	0.745 (0.743-0.746)	**0.499** (0.489-0.509)	0.857 (0.849-0.865)	0.731 (0.727-0.735)	0.789 (0.787-0.791)	0.719 (0.718-0.720)
Stacked Ensemble[§]	**0.750** (0.749-0.751)	0.493 (0.492-0.494)	0.869 (0.868-0.870)	**0.732** (0.730-0.733)	0.795 (0.794-0.796)	**0.724** (0.723-0.725)
Best	**0.750**	**0.499**	**0.876**	**0.732**	**0.796**	**0.724**
Resampling: Up						
NB	0.676 (0.674-0.678)	0.516 (0.509-0.523)	0.759 (0.753-0.764)	0.599 (0.597-0.600)	0.669 (0.667-0.671)	0.634 (0.633-0.636)
LR	0.734 (0.732-0.736)	**0.675** (0.673-0.677)	0.674 (0.671-0.677)	0.664 (0.663-0.665)	0.669 (0.668-0.670)	0.674 (0.673-0.675)
CART	0.726 (0.724-0.729)	0.620 (0.611-0.629)	0.749 (0.737-0.762)	0.652 (0.650-0.655)	0.698 (0.693-0.702)	0.683 (0.682-0.684)
XGB	0.747 (0.747-0.747)	0.625 (0.619-0.630)	0.750 (0.744-0.757)	0.656 (0.652-0.660)	0.700 (0.699-0.701)	0.686 (0.685-0.687)
RF	0.778 (0.777-0.779)	0.654 (0.653-0.656)	0.763 (0.762-0.765)	0.678 (0.676-0.679)	0.718 (0.717-0.719)	0.708 (0.707-0.709)
ADA	0.743 (0.742-0.745)	0.651 (0.648-0.653)	0.717 (0.715-0.720)	0.662 (0.660-0.663)	0.688 (0.687-0.690)	0.683 (0.681-0.685)
Stacked Ensemble[§]	**0.787** (0.784-0.790)	0.664 (0.656-0.671)	**0.771** (0.757-0.785)	**0.686** (0.683-0.688)	**0.726** (0.720-0.732)	**0.716** (0.713-0.719)
Best	**0.787**	**0.675**	**0.771**	**0.686**	**0.726**	**0.716**
Resampling: Down						
NB	0.676 (0.674-0.679)	0.515 (0.476-0.555)	**0.758** (0.742-0.774)	0.599 (0.585-0.613)	0.669 (0.666-0.671)	0.634 (0.621-0.646)
LR	0.734 (0.732-0.736)	**0.675** (0.672-0.677)	0.674 (0.671-0.677)	**0.664** (0.661-0.667)	0.669 (0.667-0.672)	0.674 (0.672-0.677)
CART	0.727 (0.725-0.729)	0.625 (0.612-0.637)	0.744 (0.731-0.757)	0.654 (0.651-0.658)	0.696 (0.677-0.685)	0.683 (0.681-0.685)
XGB	0.744 (0.743-0.746)	0.613 (0.600-0.627)	**0.758** (0.744-0.772)	0.652 (0.645-0.658)	**0.701** (0.692-0.700)	0.684 (0.683-0.685)
RF	0.749 (0.748-0.750)	0.629 (0.628-0.630)	0.750 (0.749-0.751)	0.659 (0.658-0.660)	**0.701** (0.700-0.702)	**0.688** (0.687-0.689)
ADA	0.744 (0.742-0.745)	0.646 (0.641-0.650)	0.723 (0.721-0.726)	0.661 (0.660-0.662)	0.691 (0.689-0.692)	0.684 (0.682-0.685)
Stacked Ensemble[§]	**0.750** (0.748-0.751)	0.637 (0.631-0.642)	0.742 (0.733-0.752)	0.661 (0.658-0.664)	0.699 (0.697-0.702)	**0.688** (0.686-0.690)
Best	**0.750**	**0.675**	**0.758**	**0.664**	**0.701**	**0.688**

[§]The base learners are XGB, RF, ADA, and the super learner is LR

[‡] The boldfaced numbers indicate the highest value of a predictive performance indicator

Note: NB: Naïve Bayes, LR: logistic regression, CART: classification and regression tree, XGB: extreme gradient boost, RF: random forest, ADA: adaptive boosting

C.2. Hyperparameter Tuning

Table C2. Hyperparameter search space and optimal values for ML algorithms

ML Algorithms	Hyperparameter Configuration			
	Parameters	**Ranges**	**Optimal Value**	**Possible range**
NB	Laplace	0, 0.5, 1	0	>0
LR	Iteration	10, 50, 100	50	>1
CART	CP	0, 0.0001, 0.001, 0.01, 1	0.0001	0-1
	Max depth	10, 20, 30, 50	30	>1
XGB	Gamma	1, 10, 100, 1000	10	>0
	Max depth	10, 20, 30, 50	30	>1
	Eta	0.1, 0.001, 1	1	0-1
	Min-child-weight	1, 3, 5, 10	5	>0
	Subsample	0.5, 1	1	0-1
RF	Max depth	10, 20, 30, 50	20	>1
	N-trees	10, 50, 100, 500, 1000	500	>1
ADA	Max depth	10, 20, 30, 50	30	>1
	Loss function	Logistic, exponential	Logistic	Logistic, exponential
	CP	0, 0.0001, 0.001, 0.01,0.1	0.0001	0-1

Note: NB: Naïve Bayes, LR: logistic regression, CART: classification and regression tree, XGB: extreme gradient boost, RF: random forest, ADA: adaptive boosting
CP: complexity,

C.3. Pareto Optimality

Table C3. Pareto **optimal** points (nondominated ML models) in terms of six performance measures

Algorithm	Resample technique	AUC	Recall/ Sensitivity	Specificity	Precision	F1	Accuracy
RF	**No**	0.75	0.49	0.88	0.73	0.8	0.73
XGB	**No**	0.75	0.49	0.88	0.73	0.8	0.73
ADA	**No**	0.75	0.5	0.86	0.73	0.8	0.73
RF	**Down-sample**	0.75	0.63	0.75	0.77	0.77	0.7
RF	**Up-sample**	0.75	0.62	0.77	0.76	0.76	0.71
Stacked Ensemble§	**Down-sample**	0.75	0.64	0.74	0.77	0.75	0.7
ADA	**Up-sample**	0.75	0.65	0.72	0.77	0.74	0.69
ADA	**Down-sample**	0.75	0.65	0.72	0.77	0.74	0.69
Stacked Ensemble§	**Up-sample**	0.74	0.59	0.78	0.75	0.77	0.71
LR	**Up-sample**	0.73	0.68	0.67	0.77	0.72	0.68

§ The base learners are XGB, RF, ADA, and the super learner is LR
‡ The gray lines indicate non-dominated ML models in terms of AUC (function of sensitivity and specificity), F1 (function of recall and precision), and accuracy
Note: LR: logistic regression, XGB: extreme gradient boost, RF: random forest, ADA: adaptive boosting

Figure C1. Pareto frontier in terms of AUC, F1 measure, and accuracy

C.4. Calibration

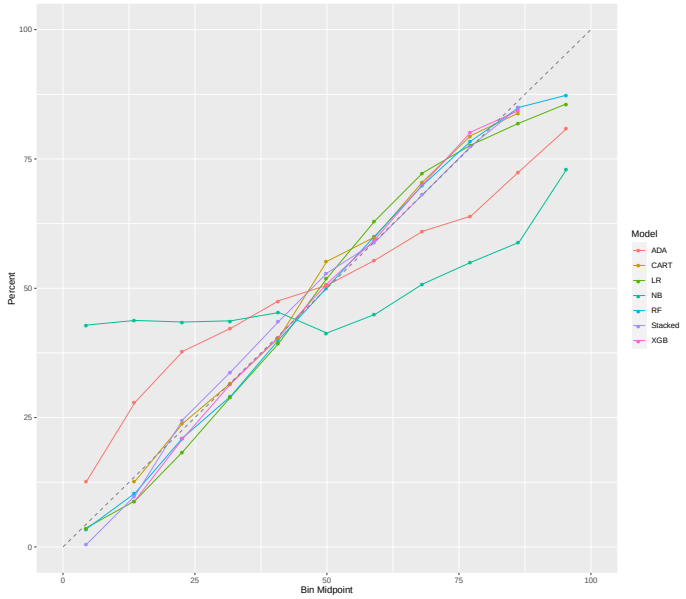


Figure C2. Comparing the Calibration of Constructed Models
Note: NB: Naïve Bayes, LR: logistic regression, CART: classification and regression tree, XGB: extreme gradient boost, RF: random forest, ADA: adaptive boosting

Appendix D. Data Eligibility Flow Diagram

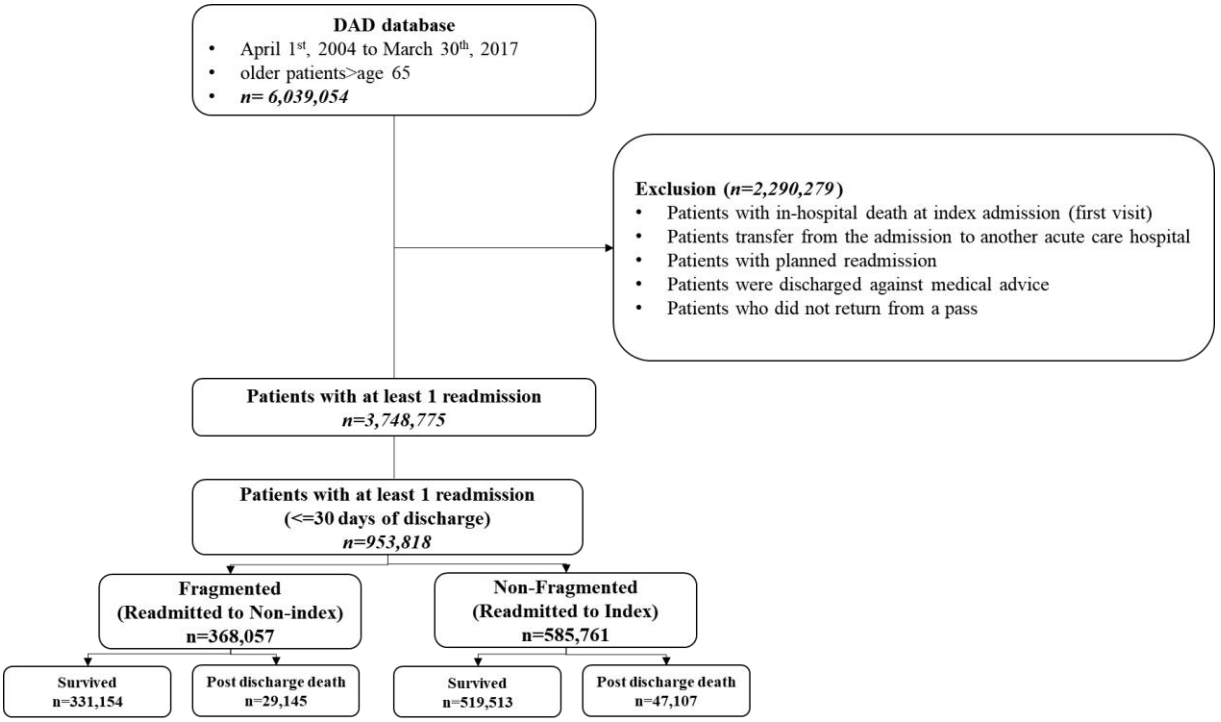
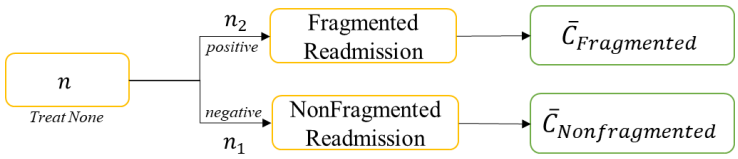


Figure D1. Data eligibility flow diagram

Appendix E. Decision Analysis

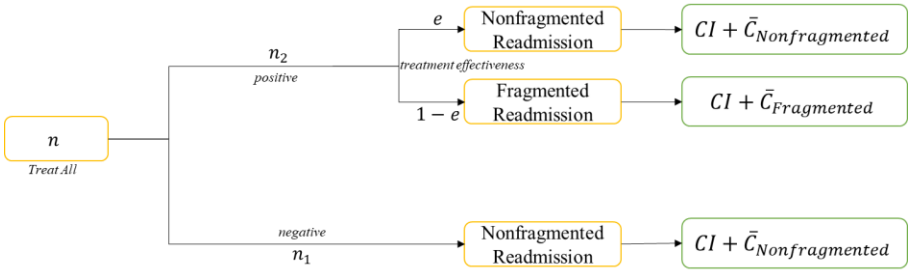
Appendix E1. Decision trees associated with each strategy

a) treat none



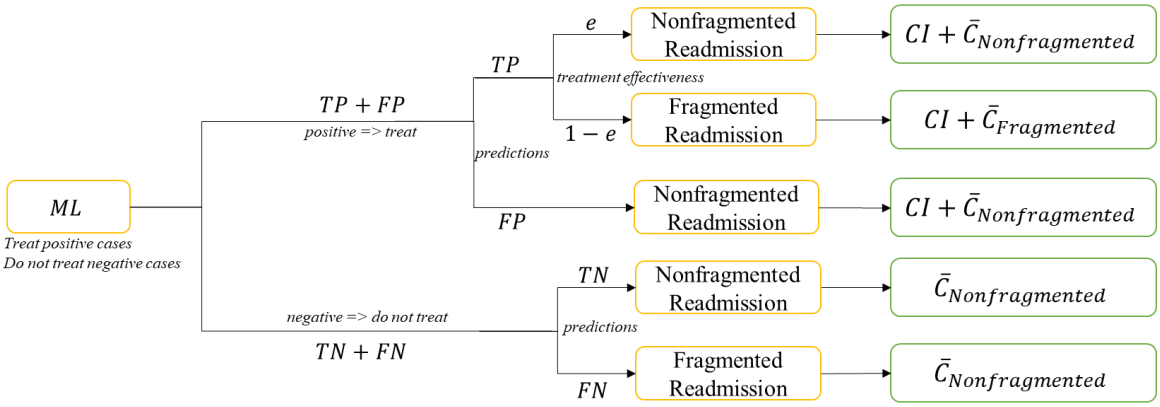
$$TC_{none} = n_2\bar{C}_{Fragmented} + n_1\bar{C}_{Nonfragmented}$$

b) treat all



$$TC_{all} = n_2 e\left(\bar{C}_{Nonfragmented} + CI\right) + n_2(1-e)\left(\bar{C}_{Fragmented} + CI\right) + n_1\left(\bar{C}_{Nonfragmented} + C_I\right) =$$
$$= nCI + n_2(1-e)\bar{C}_{Fragmented} + (n_2 e + n_1)\bar{C}_{Nonfragmented}$$

c) Treat positive cases predicted by ML



$$TC_{ML} = TP.e.\left(CI + \bar{C}_{Nonfragmented}\right) + TP(1-e)\left(CI + \bar{C}_{Fragmented} +\right) + FP\left(CI + \bar{C}_{Nonfragmented}\right)$$
$$+ TN.\bar{C}_{Nonfragmented} + FN.\bar{C}_{Fragmented}$$
$$= CI(TP + FP) + \bar{C}_{Fragmented}(TP(1-e) + FN) + \bar{C}_{Nonfragmented}(TP.e + FP + TN)$$

Figure E1. Decision trees associated with the cost of each strategy:
a) treat none, b) treat all, c) treat based on ML

Appendix E2. Sensitivity analysis on the event rate

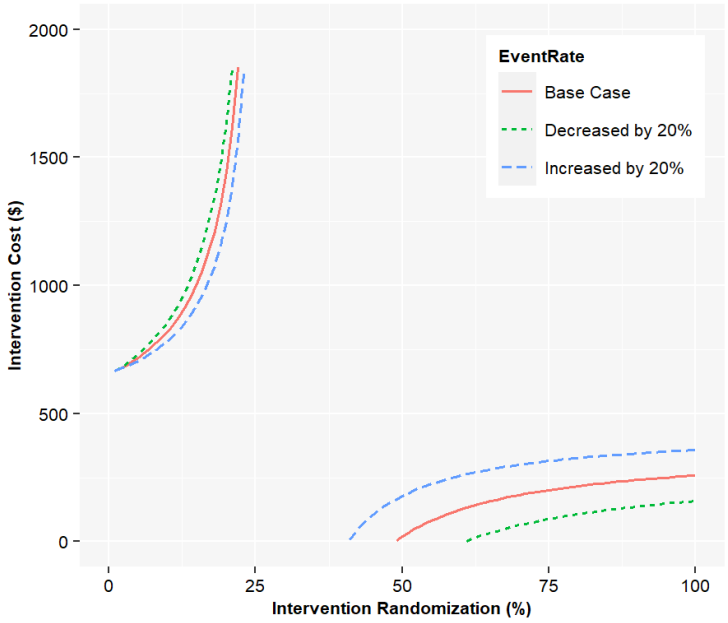


Figure E2. Optimality area of ML w.r.t the random strategies: sensitivity analysis on the event rate ($e = 0.5, P_t = 0.5$)

Appendix E3. Personalized Analysis for the Patient Complexity Groups

Table E1. Comparisons of cost and model performance across patient complexity groups

Groups	Index cost	Non-index cost	Diff	Sensitivity	Specificity	Precision
Frailty (H), Comorbidity (H)	19,470	27,895	8,425	0.36	0.89	0.68
Frailty (H), Comorbidity (L)	13,420	16,189	2,769	0.35	0.88	0.69
Frailty (L), Comorbidity (H)	9,070	10,822	1,752	0.46	0.87	0.68
Frailty (L), Comorbidity (L)	6,928	7,595	667	0.53	0.86	0.72
Overall Population	9,448	11,344	1,896	0.48	0.87	0.70

H: high level, L: low level, diff: non-index cost- index cost

Appendix F. Fairness Analysis

Appendix F1. Fairness Definitions and Formula

Table F1. Fairness Notions (formula and descriptions)

Fairness Definition	Formula	Definition
Statistical parity	(TP+FP)/n	Equal probability of being assigned to the fragmented readmission for both protected and nonprotected groups.
Equal opportunity	TP / (TP + FN)	Equal probability of being correctly assigned fragmented readmission to actually fragmented readmission for both protected and nonprotected groups (equal sensitivity).
Predictive parity	TP / (TP + FP)	Equal probability of correct fragmented readmission predictions for both protected and nonprotected groups (equal precision).
Predictive equality	TN / (TN + FP)	Equal probability of correct nonfragmented readmission predictions for both protected and nonprotected groups (equal specificity)
Accuracy parity	(TP + TN)/n	Equal probability of being correctly assigned nonfragmented readmission to actual nonfragmented readmission AND being correctly assigned fragmented readmission to actually fragmented readmission for both protected and nonprotected groups.
FNR parity	FN / (TP + FN)	Equal probability of being incorrectly assigned nonfragmented readmission to actually fragmented readmission for both protected and nonprotected groups.
FPR parity	FP / (TN + FP)	Equal probability of being incorrectly assigned fragmented readmission to actual nonfragmented readmission for both protected and nonprotected groups.
NPV parity	TN / (TN + FN)	Equal probability of being correctly assigned nonfragmented readmission to actual nonfragmented readmission for both protected and nonprotected groups.
AUC parity	-	Equal ROC AUC values for both protected and nonprotected groups.
MCC parity	(TP×TN-FP×FN)/√((TP+FP)×(TP+FN)×(TN+FP)×(TN+FN))	Equal MCC scores, which summarize both fragmented and nonfragmented cases with the same weight of importance into a single informative metric.
Calibration	-	The notion of calibration is similar to predictive parity, considering the fraction of correct positive predictions for any probability threshold. Patients in both the protected and unprotected groups should have an equal probability of truly belonging to a positive class for any probability threshold.

Appendix F2. Fairness Measurements

Table F2. Predictive equality (specificity) in different protected subpopulations and thresholds

	$(Predictive\ Equality \mid P_t, PG)$ §										
Protected Groups (PG):	$P_t{}^{*}$ ‡	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
Female	0.5	0.00	0.08	0.62	0.79	**0.88**	0.93	0.96	0.99	1	1
Male	0.5	0.00	0.08	0.61	0.80	**0.87**	0.94	0.97	0.99	1	1
Rural	0.6	0.00	0.09	0.38	0.55	0.68	0.81	**0.88**	0.98	1	1
Urban	0.4	0.00	0.08	0.66	**0.86**	0.91	0.95	0.98	0.99	1	1
Most Marginalized	0.5	0.00	0.01	0.66	0.87	**0.87**	0.97	0.98	0.99	1	1
Low-moderate Marginalized	0.5	0.00	0.10	0.61	0.78	**0.87**	0.92	0.96	0.99	1	1

§ : $(Predictive\ Equality \mid P_t, PG) = (Specificity \mid P_t, PG)$
‡: *the best threshold*

Table F3. Equal opportunity (sensitivity) in different protected subpopulations and thresholds

	$(Equal\ Opportunity \mid P_t, PG)$ §										
Protected Groups (PG):	$P_t{}^{*}$ ‡	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
Female	0.5	1	0.98	0.74	0.60	**0.48**	0.39	0.30	0.15	0.00	0.00
Male	0.5	1	0.98	0.73	0.56	**0.48**	0.34	0.25	0.12	0.00	0.00
Rural	0.6	1	0.99	0.92	0.84	0.74	0.60	**0.48**	0.21	0.00	0.00
Urban	0.3	1	0.98	0.66	**0.49**	0.38	0.28	0.20	0.10	0.00	0.00
Most Marginalized	0.5	1	1	0.61	0.38	**0.48**	0.22	0.16	0.11	0.00	0.00
Low-moderate Marginalized	0.5	1	0.98	0.76	0.61	**0.48**	0.39	0.29	0.14	0.00	0.00

§ $(Equal\ Opportunity \mid P_t, PG) = (Sensitivity \mid P_t, PG)$; ‡ *the best threshold*

Table F4 shows the calibration scores (probability of truly belonging to the fragmented readmission class) for the protected groups. The scores are slightly different for men and women in each bin. The same pattern can be observed for marginalized groups. In contrast, the scores are quite different in lower values of the threshold ($P_t < 0.5$) for rural and urban patients but become closer for thresholds greater than 0.5. Thus, we could conclude that our model fully satisfies this fairness definition concerning sex and marginalization but only partially satisfies it for the other protected group.

Table F4. Comparison of the calibration scores for different protected subpopulations and thresholds

	$P(fragmented\ readmission \mid P_t, PG)$ §									
Protected Groups (PG):	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
Female	0.40	0.42	0.57	0.66	0.71	0.78	0.83	0.92	NA	NA
Male	0.37	0.38	0.52	0.62	0.70	0.76	0.81	0.89	NA	NA
Rural	0.53	0.55	0.63	0.68	0.72	0.78	0.82	0.91	NA	NA
Urban	0.35	0.36	0.51	0.62	0.69	0.76	0.82	0.90	NA	NA
Most Marginalized	0.36	0.36	0.50	0.63	0.71	0.78	0.85	0.90	NA	NA
Low-moderate Marginalized	0.39	0.41	0.55	0.64	0.70	0.77	0.82	0.90	NA	NA

§ $P(fragmented\ readmission \mid P_t, PG) = (precision \mid P_t, PG) = (Predictive\ Parity \mid P_t, PG)$

Appendix F3. Cost and Fairness measurements

Table F5. Cost and fairness measures of marginalization attribute

Groups	**Index cost**	**Non-index cost**	**Diff**	**EO before**	**PE before**	**EO after**	**PE after**
Low-moderate marginalized	9,276	11,062	1,786	0.48	0.87	0.48	0.87
Most marginalized	10,443	12,745	2,302	0.48	0.87	0.48	0.87
Overall population	9,448	11,344	1,896	0.48	0.87	0.48	0.87

Diff= non-index cost-Index cost; EO: Equal opportunity; PE: Predictive Equality

Table F6. Cost and fairness measures of sex attribute

Groups	**Index cost**	**Non-index cost**	**Diff**	**EO before**	**PE before**	**EO after**	**PE after**
Female	9,271	10,937	1,665	0.48	0.87	0.48	0.87
Male	9,673	11,647	1,973	0.48	0.88	0.48	0.88
Overall population	9,448	11,344	1,896	0.48	0.87	0.48	0.87

Diff= non-index cost-Index cost; EO: Equal opportunity; PE: Predictive Equality

Table F7. Cost and fairness measures of residency attribute

Groups	**Index cost**	**Non-index cost**	**Diff**	**EO before**	**PE before**	**EO after**	**PE after**
Rural	8,093	9,867	1,774	0.74	0.68	0.48	0.88
Urban	9,731	11,889	2,158	0.38	0.91	0.49	0.86
Overall population	9,448	11,344	1,896	0.48	0.87	0.48	0.87

Diff= non-index cost-Index cost; EO: Equal opportunity; PE: Predictive Equality

Appendix F4. Difference between Event Rate and Equal Opportunity

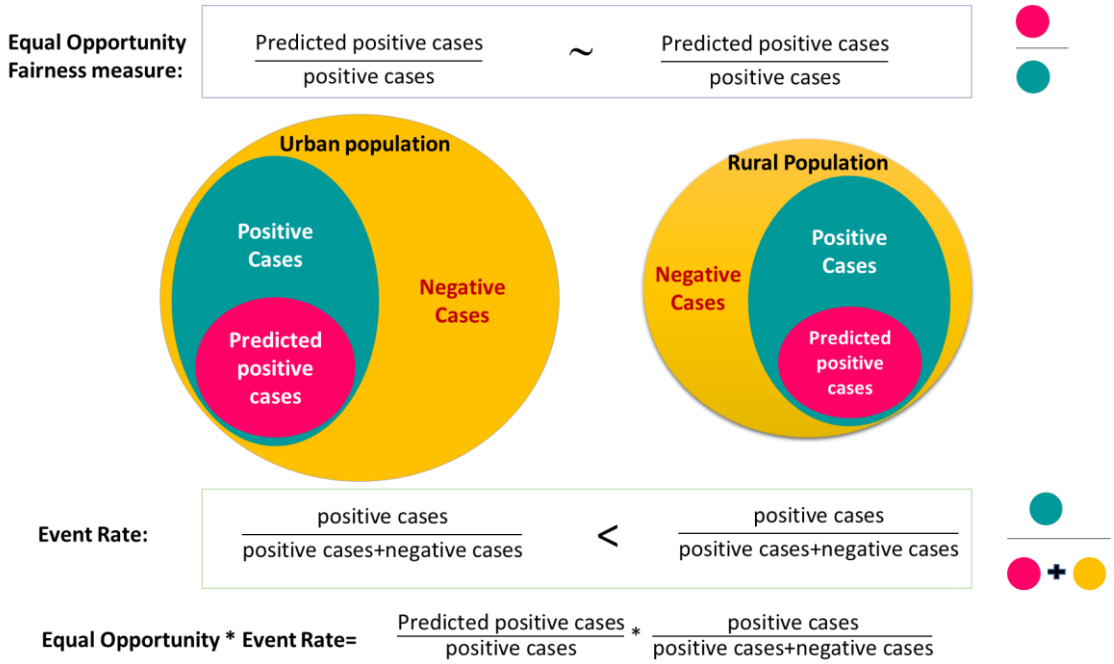


Figure F1. The difference between equal opportunity and event rate between rural and urban residents

Appendix F5. Subgroup and Intersectional Groups Fairness Analysis

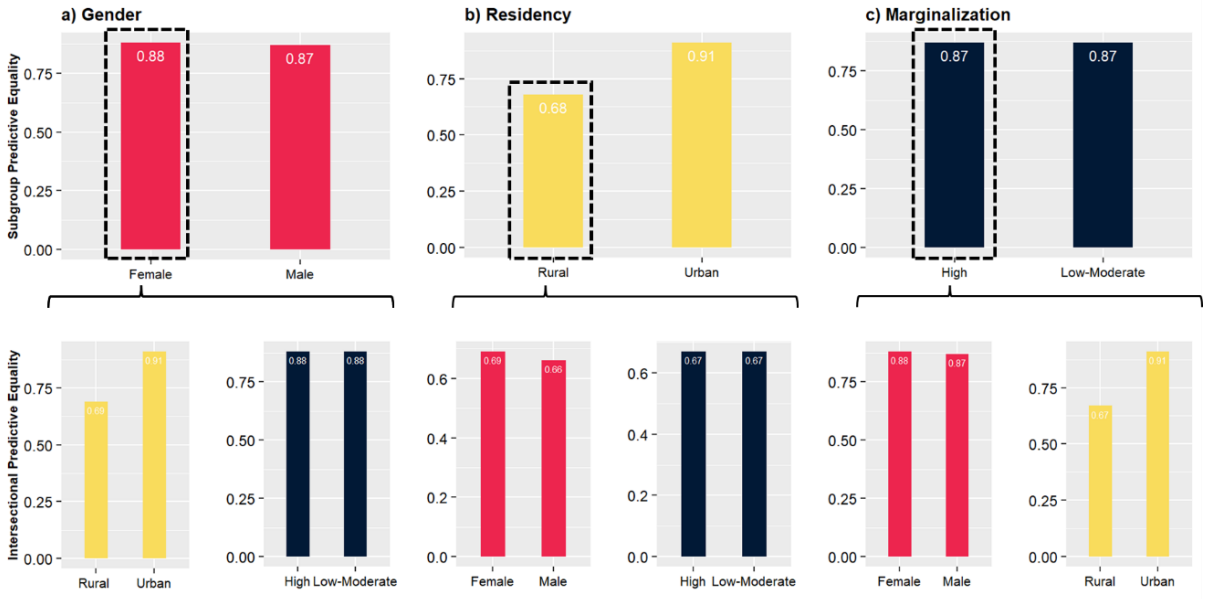


Figure F2: Analysis of predictive equality across subgroup and intersectional groups

Appendix G. Real-World Application

Appendix G1. Standardization of program costs

Table G1. The program costs extracted from studies and used to standardize costs to 2020 CAD

Program & Study	Costs extracted from studies			Currency conversion		Inflation-adjusted	e
	Currency	Year	Cost	Exchange	C_I^p(CAD) ‡	C_I^p (CAD) §	%
Odeh et al.							
90-day Phone call	UK £	2019	35.99	1.719	61.899	63.23	24.0
Wong et al.							
Home visit	Hong Kong $	2012	997	0.173	172.481	195.53	24.9
Phone call	Hong Kong $	2012	451	0.173	78.023	88.45	24.2
Combined	Hong Kong $	2012	1,448	0.173	250.504	283.98	25.5
Gardner et al.							
Home visit & Phone call	USD $	2011	298	1.342	399.767	465.10	26.8
Ornstein et al.							
HBPC program	USD $	2008	374	1.342	501.721	614.41	27.9
Saleh et al							
PDCT program	USD $	2009	946	1.342	1,269.059	1,532.24	34.3

HBPC: home-based primary care; PDCT: post-discharge care transition; C_I^p: intervention/program cost; e: effectiveness rate

‡ *Converted cost (per patient) after applying the exchange rate*

§ *Cost of the program per patient*

Appendix G2. The trend of the ML cost-saving (Sample size:1 million)

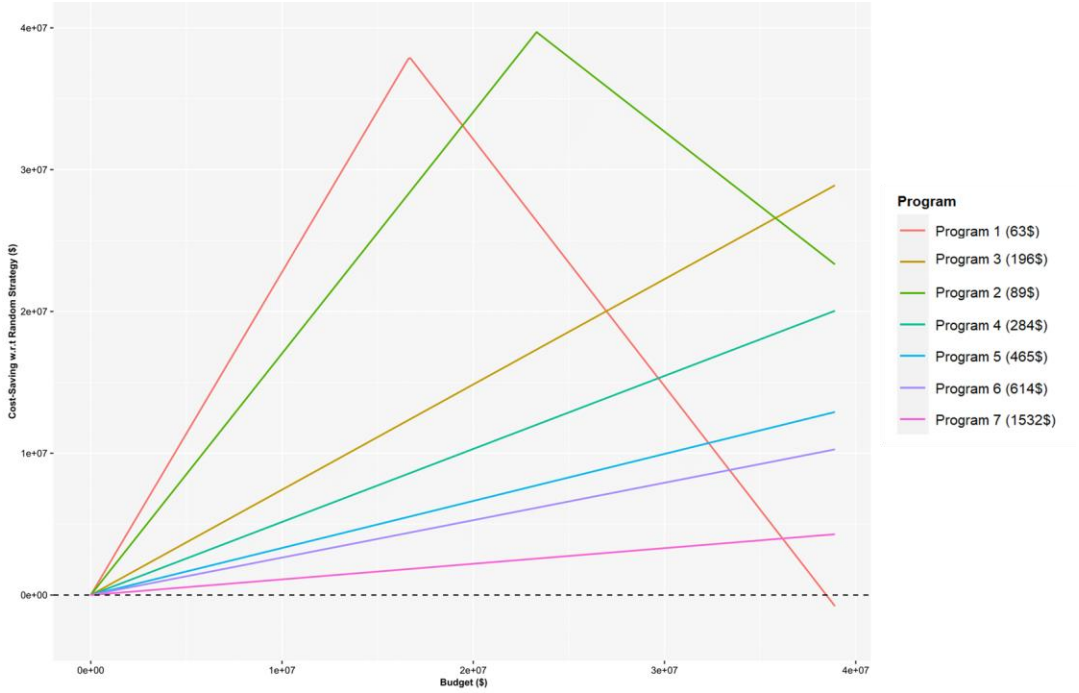


Figure G1. The trend of the ML cost-saving by increasing the budget for different programs (Sample size:1M)

Appendix H: Robustness Check on the Cause of COC

Table H1. ML cost saving relative to the random strategy for different programs (After removing cancer-related observations)

Programs	C_I ($)	e (%)	ML budget (M$)	Random strategy (M$)	ML strategy (M$)	Absolute Saving (M$)	Relative Saving (%)
Odeh et al.							
90-day Phone call	63.23	24.0	1.57	948.02	944.33	3.68	0.385
Wong et al.							
Phone call	88.45	24.2	2.19	951.13	947.31	3.82	0.401
Home visit	195.53	24.9	4.85	948.61	944.90	3.71	0.391
Combined	283.98	25.5	7.04	953.21	949.30	3.91	0.410
Gardner et al.							
Home visit & Phone call	465.10	26.8	11.53	957.46	953.35	4.11	0.429
Ornstein et al.							
HBPC program	614.41	27.9	15.23	960.96	956.68	4.28	0.445
Saleh et al							
PDCT program	1,532.24	34.3	37.98	982.54	977.28	5.26	0.530

HBPC: home-based primary care; PDCT: post-discharge care transition; C_I: program cost per patient; e: effectiveness rate

Chapter 5
Conclusion

Summary and Implications

COC is about delivering seamless patient services, continuous caring relationships, and information sharing between patients and care providers (Gulliford et al. 2006). A disruption in COC (referred to as CF) is one of the main sources of inefficiency in the healthcare system, and such events lead to increased healthcare costs (Rosenberg and Zulman 2020) and reduced quality of care (Hirji et al. 2020). This issue is more serious among older patients with medically complex needs (Brooke 2020, Meijboom et al. 2010), whose prevalence in Canada has tripled in size over the last 40 years and is expected to grow by 68% in the next 20 years (CIHI 2017). This dissertation, which includes three essays, outlined a program of research aiming to design data-driven analytic frameworks to address the existing methodological and empirical gaps regarding COC among older adults in the Canadian healthcare system. To this end, several data-driven modeling frameworks were designed based on statistical and artificial intelligence techniques—specifically, ML—and decision-analysis techniques using rich data collected over a decade. I used these frameworks to provide important insights into decision and policy making at both the service and systems levels that can potentially improve COC and mitigate the negative impact of CF. Three distinct yet related studies, as described in Chapters 2, 3, and 4, were conducted to pursue these goals.

In Chapter 2, we proposed a data-driven predictive analytic framework that leveraged big data and ML methodologies to examine the predictability and prognostication of patient complexity, in terms of multimorbidity, among older adults. This was particularly relevant to delayed-discharge patients. We first examined the predictability of three common multimorbidity indices—namely,

the CDCI, the ECI, and the FCI—using ML. We then assessed the prognostic power of these indices for predicting 30-day readmission and mortality. Our findings highlight the feasibility and utility of predicting multimorbidity status using ML algorithms, resulting in the early detection of patients at risk of mortality and readmission. This can support proactive triage and decision-making about staffing and resource allocation, with the goal of optimizing patient outcomes and facilitating an upstream and informed discharge process via prioritizing complex patients for discharge and providing patient-centered care, which is aligned with improving COC. Hence, the main contributions of the first essay can be summarized as follows:

1- We tested the feasibility and performance of data-driven ML-based models to predict patient complexity in terms of multimorbidity.
2- We also investigated the prognostication power of patient complexity.
3- Within a single study, we linked two results of (1) and (2), providing important clinical insights regarding the relationship between predictability and prognostication of patient complexity.
4- We highlighted the importance and feasibility of early prediction of multimorbidity status for optimizing discharge planning and transition of care programs in the context of older delayed discharge patients.

In Chapter 3, we propose a data-driven predictive analytical framework that leverages big data and statistical methodologies to assess the effects of patient complexity on a series of patient outcomes among older patients with delayed discharge. We examined the coexisting effects of multimorbidity and frailty as two components of patient complexity, measured using the ECI (Azzalini et al. 2019, Elixhauser et al. 1998) and the HFRS (Gilbert et al. 2018), respectively, on 30-day mortality and two means of hospital readmission (via the ED or directly) within 30 days after discharge. This study highlights the importance of considering coexisting multimorbidity and frailty, in addition to several other patient-specific factors (e.g., sex, residency, and marginalization status), to better understand the complex needs of older delayed-discharge patients and to inform

discharge policies by prioritizing patients at risk for adverse outcomes. Advanced knowledge of these factors could support proactive, informed, and equitable discharge planning and clinical decision-making, given the greater risk of delayed discharge in older adults with complex conditions (Bhatia et al. 2020). We speculate that our study's insights about the pre-and post-discharge policies would still apply to older adults without delayed discharge. For instance, regardless of a delayed-discharge designation, patient complexity could be assessed and reported to the clinical and managerial teams during hospitalization to inform the intensity and type of post-discharge care for the patients (even those discharged without delay). Hence, the main contributions of the second essay can be summarized as follows:

1- We examined the joint effects of two key dimensions of patient complexity—that is, multimorbidity and frailty—on patient outcomes.
2- We investigated the dependency of this effect on important policy factors, such as socio-economic and demographic factors, including sex, residency, and marginalization.
3- We proposed policymaking insights regarding care transitions among older delayed discharge patients.
4- We proposed managerial insights regarding better decision-making with optimal allocation of resources for older delayed discharge patients.

In Chapter 4, we proposed a data-driven predictive–prescriptive analytics framework that leveraged big data and ML methodologies. The aim of leveraging these methodologies was to drive optimal intervention screening policies for preventing CF while addressing disparities in the decision-making process. To this end, we accomplished the following objectives: a) we developed a competitive ML-based prediction model to identify patients at risk of fragmented readmission, b) we illustrated and investigated how ML predictions can be used for targeted interventions in real clinical practices via extensive comparison with random intervention strategies, and c) we examined the fairness implications of the developed ML-based decision-making framework to

ensure parity among protected groups. Our findings indicate that our proposed ML-based strategy outperforms existing clinical random strategies and brings significant value to managing CF. It mitigates discriminatory decisions and offers significant financial savings compared with existing strategies used in clinical practice. Our proposed framework supports decision-making and resource planning toward a targeted allocation at the systems level, and it informs actions that affect patient-centered care transition at the service level. It can also facilitate shared decision-making among the aging population, their families, and their care providers. Hence, the main contributions of the third essay can be summarized as follows:

A. Technical/Methodological Contribution:

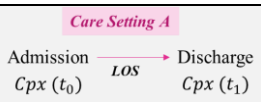
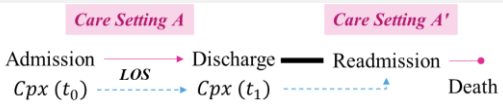
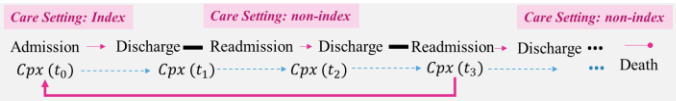
1. We developed a competitive data-driven decision analytics framework that addresses two challenges of the existing ML-based predictive-prescriptive studies, namely:
    a. how ML predictions can be explicitly used for making clinical decisions
    b. how to assess ML-based decisions' fairness (in terms of algorithmic bias) and make sure they do not lead to unfair decisions.
2. To our knowledge, we are the first to provide a decision-making framework that addresses fairness in both the predictive and prescriptive senses. To this end, we combined the notion of algorithmic bias with a need-based resource allocation philosophy to characterize fair decisions.

B. Practical Contribution:

1. We provided an evidence-based clinical decision-making framework that assists both service- and system-level decision-makers in reducing care fragmentation among older adults.
    a. For evidence-based decision-making: we utilized a rich set of longitudinal data collected over a decade with approximately 1 million unique observations to develop ML-based predictive models for CF and personalized data-driven cost estimations for clinical decision-making.
2. Our proposed recommendations can be tailored to patient attributes, providing a personalized recommendation framework.

Table 1 summarizes the proposed studies regarding their contextual and analytical characteristics.

Table 1 Summary of three proposed studies

	Context	Analytics Characteristics
Essay-1 Chapter 2		One care setting, One transition Level of analysis: Micro Level of DM: Service-level DM
Essay-2 Chapter 3		Two settings (same/different) Level of analysis: Micro Level of DM: Service-level DM
Essay-3 Chapter 4		Multiple settings (different) Level of analysis: Macro, Micro Level of DM: System-level & Service-level DM

**note* *DM: Decision Making; COC: Continuity of Care; Cpx: Complexity; LOS: Length of Stay*

Limitations

The study, consisting of the three essays, is not free from limitations. First, all three essays were retrospective, so we had limited control over data collection. It is especially true about frailty and multimorbidity, which are preferred to be measured prospectively. Second, we had limited access to some other confounders and factors driving care fragmentation, including hospital-level characteristics such as hospital size and type. Third, we had also limited access to information about the systemic issues with the infrastructure including, not enough LTC beds or senior care homes, lack of staff, and lack of adequate transport between facilities. Although we have provided platforms for decision-making, and the essence of the decisions is how to manage limited resources efficiently, information about the abovementioned resource availability can potentially improve the process of decision-making. Fourth, using modern ML algorithms such as deep learning can potentially improve the performance of our predictive models. However, because of computational limitations, I could not take advantage of them.

Further Work

In this dissertation, particularly in the second essay, we conducted a prognostication analysis to assess the association between patient complexity—encoded through frailty and

multimorbidity—and the patient-important outcomes. Future research should be devoted to predictive analytics, investigating the feasibility of predicting patient complexity, defined as a function of both multimorbidity and frailty status. We also mainly focused on two components of COC—namely, seamless services and informational continuity. Our framework can be extended to include other aspects of care continuity, such as managerial COC and continuity of physician care. It can be used to investigate how continuity among physicians and healthcare institutions can be predicted using ML algorithms. This work has also paved the way for developing frameworks for other important decision-making problems in health care, such as preventive interventions for reducing the number of patient transfers from nursing homes/long-term care facilities to acute care settings, which is one of the significant challenges among aging people in Canada (Nemiroff et al. 2019). For the sake of generalization, we examined a heterogeneous cohort of patients with various disease types. Although we adjusted on the patient complexity (multimorbidity and frailty status), future studies are needed to investigate and validate our findings on certain diseases, such as cancer or heart failure. Finally, our research is retrospective; a further prospective study is warranted to validate our framework in real-world programs and clinical settings.

References

Azzalini L, Chabot-Blanchet M, Southern DA, Nozza A, Wilton SB, Graham MM, Gravel GM, Bluteau JP, Rouleau JL, Guertin MC (2019) A disease-specific comorbidity index for predicting mortality in patients admitted to hospital with a cardiac condition. *CMAJ* 191(11):E299–E307.

Bhatia D, Peckham A, Abdelhalim R, King M, Kurdina A, Ng R, Allin S, Marchildon G (2020) *Alternate Level of Care and Delayed Discharge: Lessons Learned from Abroad* (North American Observatory on Health Systems and Policies. Rapid Review (No. 22), Toronto).

Brooke BS (2020) Transitioning Care at Discharge. *Principles and Practice of Geriatric Surgery*:485–506.

CIHI (2017) Infographic: Canada's seniors population outlook: Uncharted territory.

Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. *Med Care* 36(1):8–27.

Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, Arora S, Street A, Parker S, Roberts HC (2018) Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *The Lancet* 391(10132):1775–1782.

Gulliford M, Naithani S, Morgan M (2006) What is' continuity of care'? *Journal of health services research & policy* 11(4):248–250.

Hirji SA, Percy ED, Zogg CK, Vaduganathan M, Kiehm S, Pelletier M, Kaneko T (2020) Thirty-Day Nonindex Readmissions and Clinical Outcomes After Cardiac Surgery. *The Annals of Thoracic Surgery*.

Meijboom BR, Bakx SJ, Westert GP (2010) Continuity in health care: lessons from supply chain management. *The International journal of health planning and management* 25(4):304–317.

Nemiroff L, Marshall EG, Jensen JL, Clarke B, Andrew MK (2019) Adherence to "No Transfer to Hospital" Advance Directives Among Nursing Home Residents. *Journal of the American Medical Directors Association* 20(11):1373–1381.

Rosenberg NA, Zulman DM (2020) Measures of care fragmentation: Mathematical insights from population genetics. *Health Services Research* 55(2):318–327.