Reducing Heart Failure Admissions through Improved Care Systems and Processes

by

Mariam A. Al-Meer
B.S., Chemical Engineering, Texas A&M University at Qatar, 2012

Submitted to the MIT Sloan School of Management and the Department of Mechanical Engineering in partial fulfillment of the requirements for the degrees of
Master of Business Administration
and
Master of Science in Mechanical Engineering
in conjunction with the Leaders for Global Operations Program at the
MASSACHUSETTS INTITUTE OF TECHNOLOGY
June 2017

Author……
MIT Sloan School of Management and the Department of Mechanical Engineering
May 12, 2017

Certified by……
Retsef Levi, Thesis Supervisor
J. Spencer Standish (1945) Professor of Operations Management,
MIT Sloan School of Management

Certified by……
Brian W. Anthony, Thesis Supervisor
Principal Research Scientist, Director of the Master of Engineering in Manufacturing Program,
Department of Mechanical Engineering

Approved by……..……
Maura Herson
Director, MBA Program, MIT Sloan School of Management

Approved by……..……
Rohan Abeyaratne
Chairman, Committee on Graduate Students, Department of Mechanical Engineering

This page intentionally left blank

Reducing Heart Failure Admissions through Improved Care Systems and Processes

by

Mariam A. Al-Meer

Submitted to the MIT Sloan School of Management and the Department of Mechanical Engineering on May 12, 2017, in partial fulfillment of the requirements for the degrees of
Master of Business Administration
and
Master of Science in Mechanical Engineering

Abstract

Heart failure (HF) is a complex chronic condition that can result from any cardiac disorder that impairs the ventricle's ability to fill with or eject blood. The American Heart Association predicts that there will be about 10 million HF patients in the US by 2037, with total hospitalization costs exceeding $70 billion. This represents a considerable burden to hospitals nationwide, including the Massachusetts General Hospital (MGH) – a leading medical center that has long grappled with patient overcrowding and capacity constraints.

This thesis presents an extensive mapping of the HF care pathway at MGH, followed by the results of a detailed retrospective analysis of the general behavior of HF patients admitted to MGH. Here, we notice that the majority of HF admissions originate as self-referrals via the Emergency Department (ED) and take place on weekdays, between the hours of 9am and 6pm. Moreover, we find that about 57% of hospitalized HF patients often have no scheduled follow-up appointments with their providers in the two weeks leading up to their admissions and, similarly, about 43% have no scheduled appointments in the eight weeks post hospital discharge. These represent two critical time periods in the events of acute heart failure decompensation.

In an effort to prioritize targeted outpatient care, we propose a predictive model which aims to identify patients at greatest risk of a first hospital admission following encounters with their primary care providers and/or cardiologists in any given year. We perform logit-linear regressions on multiple prior first admissions and use predictors that, among others, include clinical risk factors, socioeconomic features and histories of prior medications. Some of the model's most significant predictors, as identified by the Akaike information criterion (AIC), include patient's age, marital status, ability to speak English, estimated average income, previous administration of loop diuretics, and the total number of medications prescribed or administered. To assess the quality of our predictions, we turn to the receiver operating characteristic (ROC) and its resulting average area under the curve (AUC) of 0.712. As the team continues to focus on developing interventions that offer better care to HF patients, the value of our model lies in its ability to prioritize patient needs for outpatient care and monitoring, and to guide the allocation of limited care resources.

Thesis Supervisor: Retsef Levi
J. Spencer Standish (1945) Professor of Operations Management
MIT Sloan School of Management

Thesis Supervisor: Brian W. Anthony
Principal Research Scientist, Director of the Master of Engineering in Manufacturing Program
Department of Mechanical Engineering

The author wishes to acknowledge the Leaders for Global Operations Program for its support of this work

Acknowledgements

The work described in this thesis would not have been possible without the generous contribution of many people. First, I would like to express my sincere gratitude to my advisor, Professor Retsef Levi, for his continued support over the last year. I am truly honored to have worked with him. His patience, optimism, and immense knowledge have not only contributed to the success of this project, but also inspired me in many ways. I thank him for his technical advice and, more importantly, for his accessibility and encouragement. I also thank my second advisor, Dr. Brian Anthony, for his input throughout the project.

I would also like to thank Dr. Peter Dunn, Bethany Daily and Cecilia Zenteno for sponsoring this project as an LGO internship. I am very grateful and proud to have had the opportunity to step into a new environment and work on a project that aims to improve the quality of people's lives. I also extend my appreciation to all members of the MGH heart failure working group for their time and dedication. I especially thank Dr. Jim Januzzi, Dr. Greg Lewis, Dr. Ryan Thompson, Dr. Kristen Livesey and Jacqui Ruckel for their clinical input and feedback, particularly in the modeling stages of this project.

I also wish to acknowledge and thank Martin Copenhaver. Martin is one of the smartest, most humble, and friendliest people that I have ever met. I am greatly indebted to him for all his help, especially when I ran into coding challenges. I am also thankful for all the talks and laughs we had while sharing office space during our summer at MGH.

I am also grateful for the LGO program and its staff. In particular, I would like to thank Patty Eames, Ted Equi and Thomas Roemer. They always had their doors open for advice and mentorship during my internship time and beyond. It is through their efforts and support that LGO is the great program it is.

The best part about LGO is the strong friendships you develop with all your classmates. In the context of my MGH internship, I want to especially thank Liz Ugarph and Jordan Hoffmann. I started my internship when they were already more than halfway through theirs and, from Day 1, they were always available to share their experiences and advice. I will always remember our weekly Friday lunches together and will always be grateful for having them around.

As I approach the end, I would like to express my heartfelt thanks to my mom, dad and four sisters. Their endless support and love are what brought me to LGO. Despite the significant time difference, they were always available for a phone call and encouraging words.

Last but not least, I want to thank my country, the State of Qatar, for generously supporting my studies at MIT. I sincerely hope that this thesis provides one form of return on investment for a country that places the education and development of its citizens at the forefront of its priorities. I dedicate this thesis to my beloved family and country.

This page intentionally left blank

Table of Contents

List of Figures

List of Tables

This page intentionally left blank

1. Introduction

1.1 Background

1.1.1 Massachusetts General Hospital

Massachusetts General Hospital (MGH) is a leading academic medical center that is located in the heart of Boston. Through its affiliation with the Harvard Medical School, it not only provides a broad range of care services to thousands of patients every year, but also trains many medical providers and scientific researchers. It offers the largest hospital-based research in the United States [1].

A 999-bed medical center, MGH admits about 48,000 inpatients annually. It performs more than 42,000 operations each year and sets the bar for quality and safety through its focus on effectiveness, patient-centeredness and timeliness [1].

Along with its dedication to deliver excellent inpatient care, MGH also focuses on improving the quality and delivery of outpatient care services at each of its primary care and specialized health centers. For example, in primary care practices alone, MGH provides services to more than 200,000 adults and children in 15 locations throughout Greater Boston. Similarly, the MGH Corrigan Minehan Heart Center brings together a team of world-class physicians and nurses, with varied heart disease specializations, to work together on offering leading treatments and preventive care for many cases of common and complex cardiac conditions [2] [3].

As one of the world's foremost academic medical centers, MGH has been consistently ranked among the top hospitals in the United States. In 2012 and 2015, it was ranked as the #1 hospital in the United States by the *U.S. News & World Report*. More recently, in 2016, it was

ranked #3 out of about 5,000 hospitals that were compared. It is the only hospital in the nation to have been ranked in all the 16 specialties it offers [4].

1.1.2 MGH-MIT Collaboration

For over a decade, MGH has collaborated with the Massachusetts Institute of Technology (MIT) Sloan School of Management on projects aimed at operational improvement. In 2011, MGH became one of the partner organizations of the MIT Leaders for Global Operations (LGO) program. This partnership has led to many system-level operational improvements in different areas of the hospital and, more generally, the overall health system. Some examples of previous improvements include patient flow within the perioperative environment, outpatient cancer center and non-oncology infusion scheduling, and primary care redesign [5] [6] [7] [8] [9] [10].

1.1.3 The MGH Capacity Task Force

To continue building on the successes of the MGH-MIT collaboration, the partnership was extended to a hospital-wide task force that was launched by the hospital's president in the final quarter of 2015. Charged with the task of alleviating capacity and access challenges, the task force is divided into the following three work groups: (i) avoidable visits and admission through the emergency department (ED), (ii) preventable readmissions, and (iii) delays related to patient misplacement and bed assignments. Each of the work groups is comprised of clinicians, hospital administrators and MIT analytics support.

1.2 Project Overview

1.2.1 Overview of the Heart Failure Condition: Classes, Stages and Progression

Heart failure can result from any structural or functional cardiac disorder that impairs the ability of the ventricle to fill with or eject blood [11]. With a 20% lifetime risk of being developed by age 40, heart failure represents a major public health problem in the United States. For example, while it was reported in 2001 that nearly 300,000 patients in this country die of heart failure as a primary or contributory cause each year, the number of deaths has more recently increased to about 610,000 despite advances in treatment [12]. Furthermore, the high rate of hospitalizations associated with heart failure puts a considerable burden to the United States healthcare system. In 2005, about 5 million patients in the United States were diagnosed with heart failure. This represented about 2% of the country's population at the time. By 2037, the number of heart failure patients is projected to increase to 10 million, or 3% of the predicted population [13]. These figures translate into more than 900,000 newly diagnosed cases of heart failure each year and total estimated annual costs of $39.2 billion dollars [11] [14].

Heart failure is primarily a disease of the elderly and approximately 80% of patients hospitalized with heart failure are older than 65 years old. In 1995, the American College of Cardiology (ACC) and the American Heart Association (AHA) first published guidelines for the evaluation and management of heart failure conditions. Since then, many developments have been observed in the pharmacological and nonpharmacological treatments of the disorder. This included a new method of classifying heart failure that focused on the evolution and progression

of the disease through four stages of heart failure[1]. These stages, along with their recommended therapies, are presented in Figure 1-1. Stage A identifies patients that have a great risk of developing heart failure but have not yet had any structural disorder of the heart. Stage B identifies those with a structural disorder of the heart but no heart failure symptoms. Stage C refers to patients with structural heart disease and past or present heart failure symptoms. Finally, Stage D groups patients with end-stage heart failure who require specialized treatment like mechanical circulatory support, continuous infusions, heart transplants or hospice care [12].

Heart failure patients, particularly those in ACC/AHA stage C or D, can also experience instances of acute decompensated heart failure (ADHF). These decompensations typically include a sudden worsening of the signs and symptoms of heart failure. In particular, they include difficulty in breathing, leg swelling and fatigue. ADHF, a common and potentially fatal cause of acute respiratory distress, often results in hospitalization and a drop in a patient's overall clinical stability [15]. This is discussed further in Section 1.2.3.

[1] It should be noted that the ACC/AHA 'Stage A to D' classification system was developed as a complement, and not replacement, of the New York Heart Association (NYHA) functional 'Class I to IV' classification system. The NYHA system is another widely used classification system that primarily gauges the severity of symptoms in patients who are in ACC/AHA stage C or D. It has long been recognized, however, that the NYHA functional classification reflects a subjective assessment by a physician and changes frequently over short durations, even though the treatments used do not differ significantly across classes [12].

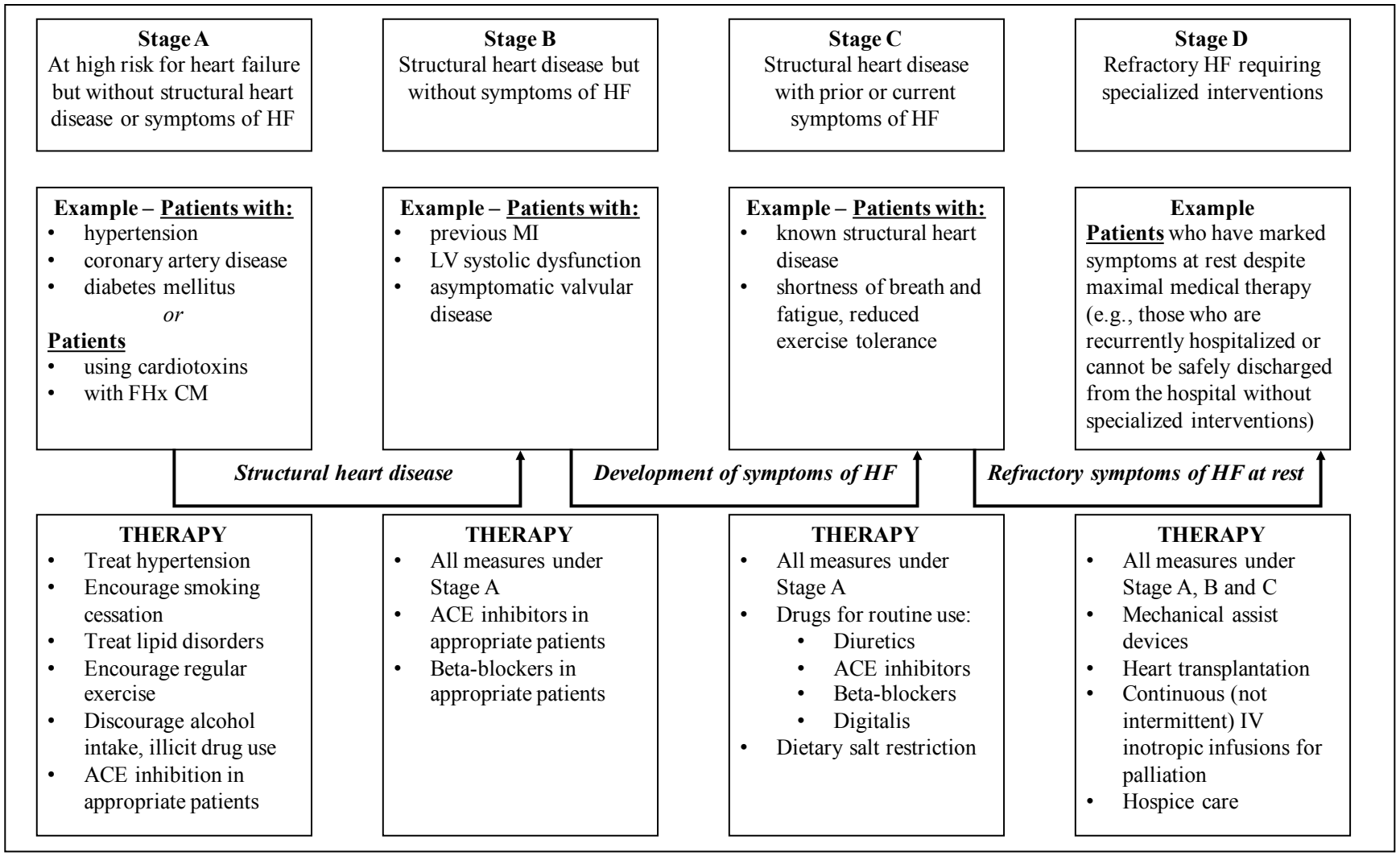


Figure 1-1: Heart Failure Stages and their Recommended Therapies; FHx CM indicates family history of cardiomyopathy; MI: myocardial infarction; LV: left ventricular; and IV: intravenous *[12]*

1.2.2 Readmissions Reduction and Rationale for Focusing on Heart Failure

In 2012, Section 3025 of the Affordable Care Act added a section to the Social Security Act around the initiation of the *Hospital Readmissions Reduction Program (HRRP)*. This program requires the Centers for Medicare and Medicaid Services (CMS) to reduce payments to inpatient prospective payment system (IPPS) hospitals[2] with excess readmissions, which is defined as an admission to a hospital within 30 days of a discharge from the same hospital, with the same condition. These measures were initially applied to the conditions of acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN) [16]. Since then, they have grown to include other conditions as well (e.g., CMS added two new readmission measures in fiscal year 2015 for chronic obstructive pulmonary disease (COPD) and elective primary total hip and/or total knee replacement) [17].

This project originated from an effort to reduce the rate of MGH 30-day readmissions among heart failure patients, which are associated with the financial penalties. However, for reasons discussed next in Section 1.2.3, the focus of the project was shifted early on from reducing readmissions to more generally lowering the overall heart failure related admissions (i.e., not limited to 30-day readmissions). The rationale for focusing on the condition of heart failure specifically is driven by three main factors. First, heart failure is associated with a high number of admissions and hospital bed days. Second, heart failure consists of a focused patient population and a relatively well-understood disease course. Third, this serves as a good candidate for improving outpatient based care management processes [11].

[2] Under the IPPS, each case is categorized into a diagnosis-related group (DRG) which, in turn, has a payment weight assigned to it based on the average resources it uses to treat Medicare patients.

1.2.3 Shifting to the Prevention of 'First-Time' Heart Failure Admissions

The focus of admission versus readmission is an attempt to tackle the problem at its root causes and design a care system that reduces the frequency of heart failure related admissions. We focus our attention on 'first-time' heart failure decompensation[3] events and their related admissions in order to reduce the likelihood that these events occur in the first place, as opposed to reducing readmissions which assume a prior occurrence by definition. Similar to the choice of focusing on heart failure as the condition of interest, the rationale for focusing on first-time heart failure admissions also consists of three parts.

First, heart failure hospitalizations are often associated with a degradation in overall patient clinical stability, and first-time hospitalizations are usually the most serious step downhill. Figure 1-2 depicts the typical progression of acute decompensated heart failure, and shows a range of possible clinical scenarios. In A, good recovery is observed after the first episode, followed by a stable period of variable length. In B, the first episode is not survived, which is represented as the patient's steep downward trajectory towards death. In C, poor recovery is observed after the first episode followed by clinical deterioration. Finally, in D, ongoing deterioration with intermittent crises and unpredictable deaths is observed [18]. One thing that seems to be significant here is how the first decompensation and, consequently, hospitalization event results in the greatest drop in clinical stability. Preventing this significant dip in stability, and the patient's entry into an active and risky state of disease progression, seems to be key.

[3] Acute decompensated heart failure (ADHF) is a sudden worsening of the signs and symptoms of heart failure, which typically includes difficulty breathing, leg swelling and fatigue. ADHF, a common and potentially fatal cause of acute respiratory distress, often results in hospitalization [15].

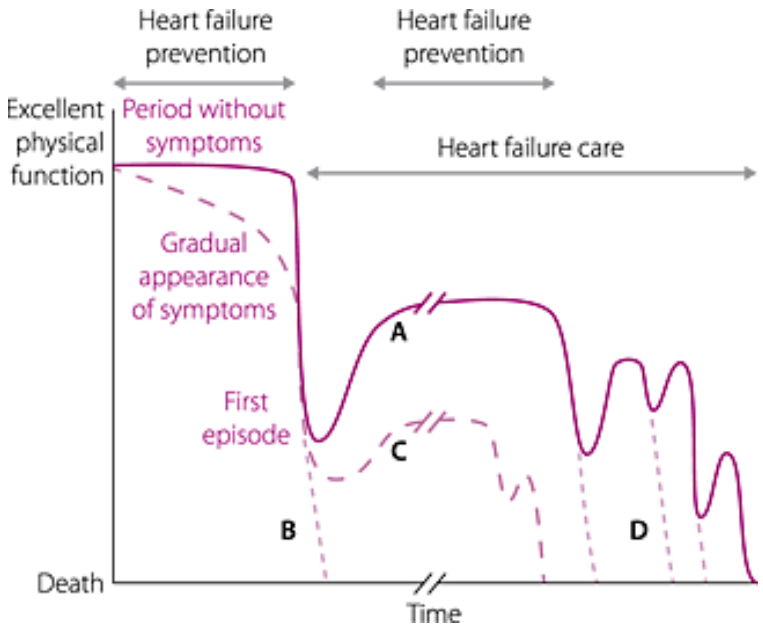


Figure 1-2: The Long-Term 'Price' of Decompensated Heart Failure *[18]*

Second, from a hospital capacity point of view, first-time heart failure admissions, despite amounting for only 5-8% of heart failure related admissions within the MGH network in a given year (see Table 5-2 in Section 5.3.1), account for about 22% of total heart failure related admission bed-days in a given year with no significant decreasing trend. This is shown in Figure 1-3.

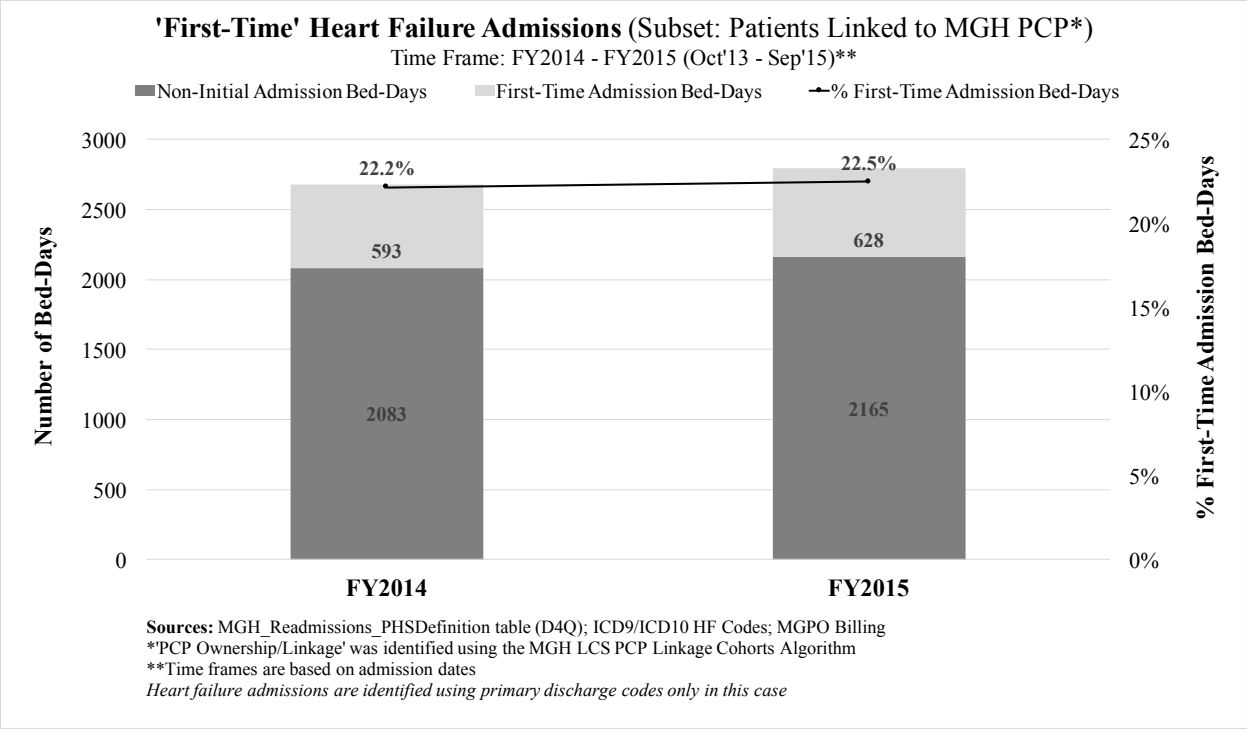


Figure 1-3: First-Time Heart Failure Admission Bed-Days as a fraction of Total Heart Failure Admission Bed-Days

Third, when the initial goal of addressing 30-day readmissions is revisited, we notice that, unlike the case of first-time heart failure related admissions, decreasing trends have already been observed for the fraction of bed-days consumed by 30-day readmissions and for 30-day readmission rates in general. This is shown in Figure 1-4 and Figure 1-5, respectively. While this does not necessarily imply that efforts around reducing heart failure related readmissions should stop, it draws our attention to the importance of focusing our work on an issue of greater concern. Specifically, we note the slightly increasing trend in fraction of bed-days consumed by first-time admissions (shown in Figure 1-3) and focus our efforts thereof.

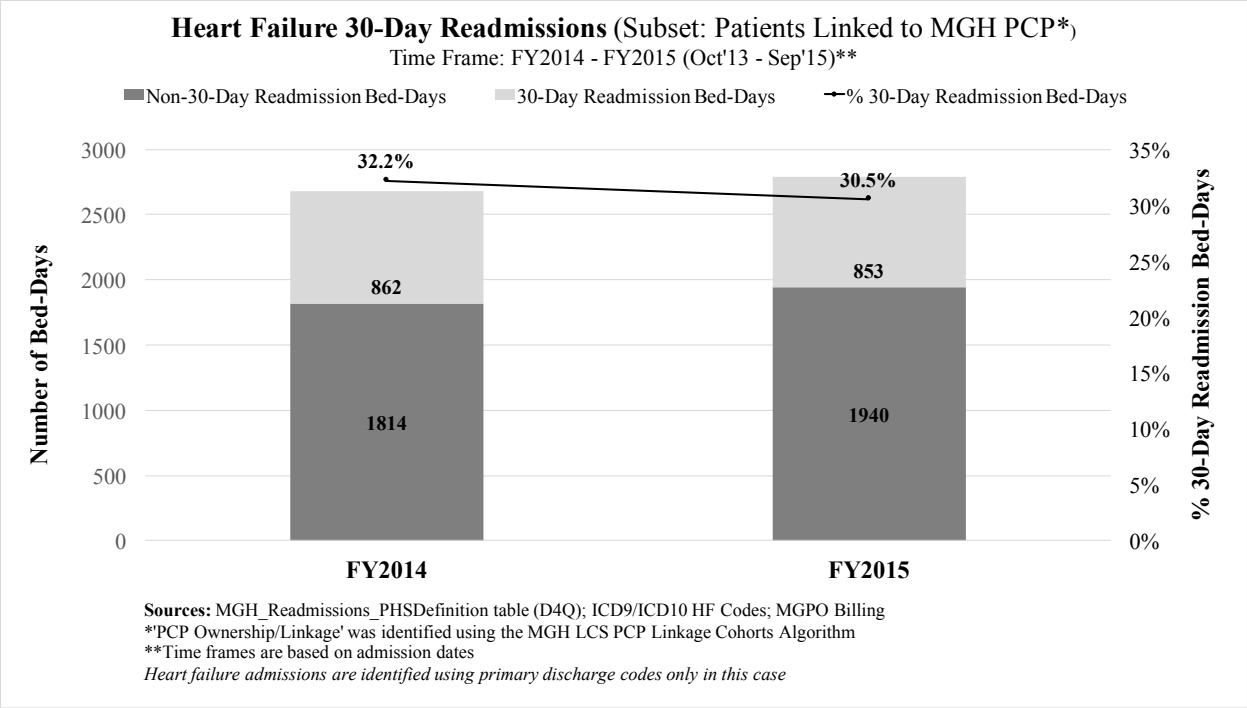


Figure 1-4: 30-Day Heart Failure Readmission Bed-Days as a fraction of Total Heart Failure Admission Bed-Days

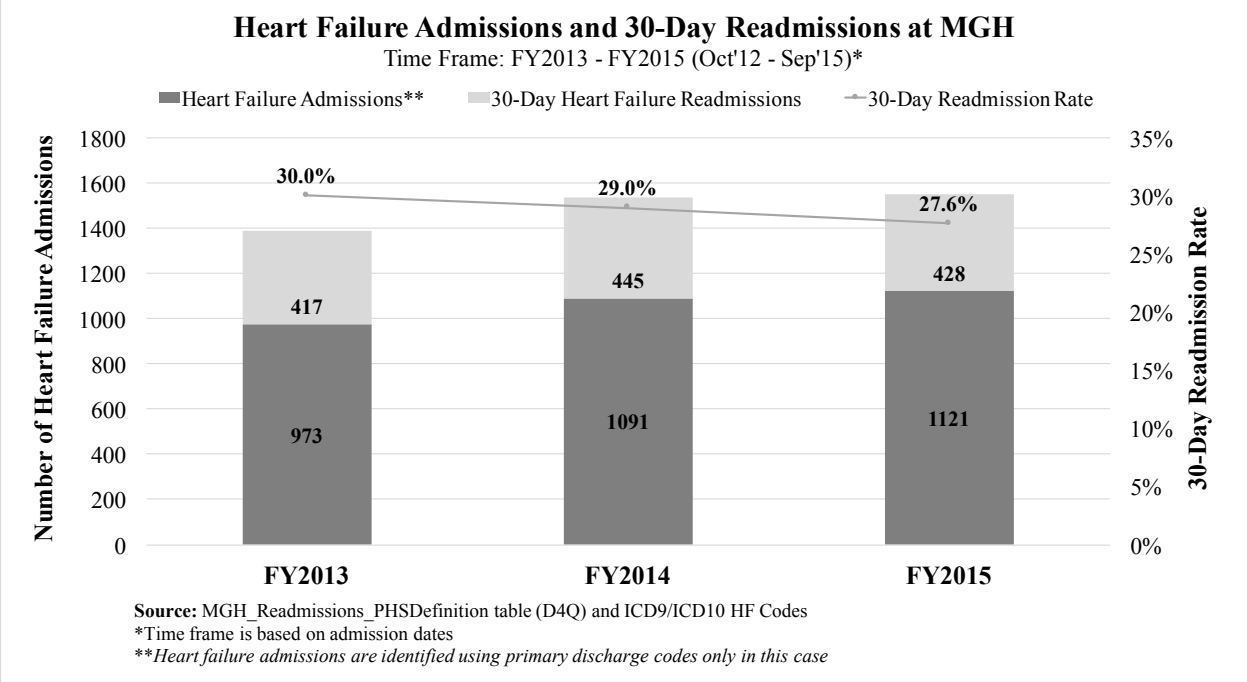


Figure 1-5: Decreasing Trend in 30-Day Heart Failure Readmission Rates, FY2013-FY2015

With these three motivating factors in mind, we focus this thesis work on understanding the current outpatient system that cares for heart failure patients and attempt to identify major areas for improvement. In addition, predictive models are developed to identify patients that are at the greatest risk of having their first heart failure admission.

1.3 Thesis Summary, Organization and Structure

Through a comprehensive mapping of the heart failure current-state care pathway, two key analyses are identified and explored in this thesis. First, we pose and validate a hypothesis that limited outpatient access is a primary cause of heart failure related admissions. Second, we explore the potential of developing a predictive model that identifies patients at greatest risk of a first-time heart failure related hospital admission.

We quantified the lack of outpatient access through a detailed analysis of heart failure patients' scheduled appointments with primary care and cardiology clinics in the two weeks prior to hospitalization and eight weeks post discharge. These represent two critical time periods in the events of acute heart failure decompensation. We found that about 57% of hospitalized heart failure patients had no scheduled follow-up appointments in the two weeks prior to their admissions. Similarly, we found that about 43% of hospitalized heart failure patients had no appointments in the eight weeks post discharge. Finally, we found that, for patients with completed appointments in the two weeks prior to their hospitalizations, the median wait time between scheduling an appointment to seeing a provider was in the range of 13-17 days.

In the second part of this thesis, we develop a model that predicts first-time heart failure related admissions with an average out-of-sample area under the receiver operating curve (AUC) of 0.712. This performance is close to other reported out-of-sample AUCs of some widely used heart failure risk-scoring models (e.g., Seattle Heart Failure Model's AUC of ~0.729). Out of a

range of 27 clinical, demographic and socioeconomic factors, we identify the 10 most significant predictive features. These include: gender; language; marital status; estimated income using zip code; lowest reported systolic blood pressure; age; time since the initial heart failure diagnosis; presence/lack of history of metoprolol administration; presence/lack of history of bumetanide administration; and the total number of medications prescribed in the year we are predicting from.

Overall, this project was approached in three phases. Initial efforts were focused at developing a general understanding of the current-state care pathway of heart failure patients at MGH. In addition, this phase included a comprehensive review of the literature for insights into the latest strategies and technologies employed in the monitoring of heart failure patients. The second phase involved a thorough quantitative analysis of the mapped heart failure care pathway. Finally, the third phase involved modeling patients' risk of a first-time admission and using the models to guide the development of targeted interventions for heart failure patients. In terms of thesis organization and structure, the chapters corresponding to each of these project phases are listed in Table 1-1.

Table 1-1: Overview of Project Phases and Corresponding Thesis Chapters

Project Phase	Details	Corresponding Thesis Chapter
1	- Map the current-state heart failure care pathway - Conduct a literature review of the latest strategies and technologies in heart failure monitoring	2 and 3
2	- Current-state data analysis: o Identify heart failure patients o Generate distributions of patients' sources and times of arrival o Quantify patients' access to outpatient resources	4
3	- Model patient risk for a first-time heart failure admission - Identify areas for future research and model improvement - Propose operational recommendations using the risk model as a guide	5 and 6

2. Literature Review

2.1 Types of Heart Failure and Advances in Drug Treatments

The heart is the muscle that pumps blood filled with oxygen and nutrients through the blood vessels to the body tissues. It is the hardest working muscle in the human body and is made up of four chambers (two atria and two ventricles) and many blood vessels (network of arteries and veins). The atria receive blood coming back to the heart through veins, and the ventricles pump the blood out of the heart through arteries. In the left side of the heart, the blood is oxygenated and ready for pumping to the rest of the body through the left atrium and ventricle. In the right side, "used" or deoxygenated blood is returning to the lungs for oxygen replenishment and this takes place through the heart's right atrium and ventricle. A cross-section of the heart is shown in Figure 2-1 [19].

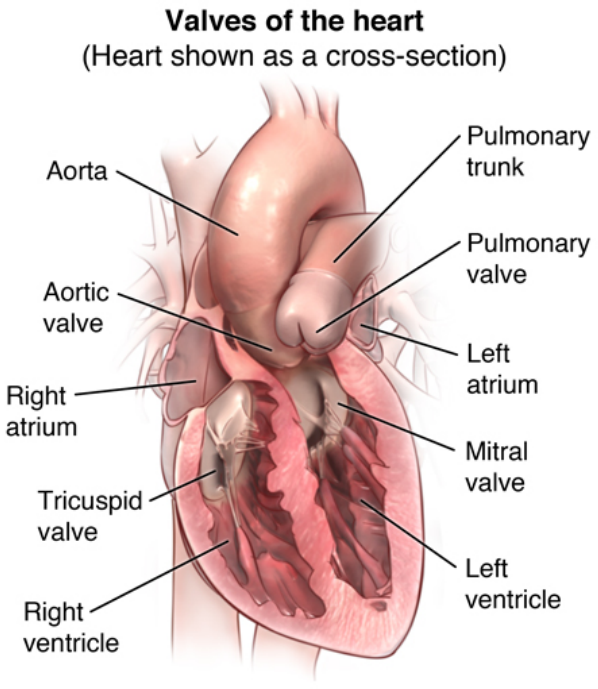


Figure 2-1: Basic Anatomy of the Heart *[19]*

Heart failure is characterized by the heart's inability to pump an adequate supply of blood. Without sufficient blood flow, all major body functions are disrupted [20]. Heart failure patients can be divided into two specific groups: (i) left-sided or left ventricular (LV) heart failure and (ii) right-sided or right ventricular heart failure. In the former, an impairment in the pumping action of the heart's left ventricle limits the supply of oxygen-rich blood as it travels from the lungs to the left atrium, then on to the left ventricle, which pumps it to the rest of the body. Left-sided heart failure can be further split into two categories. In the first case, systolic failure occurs when the left ventricle loses its ability to contract normally. This then limits the force with which oxygenated blood is pushed to the rest of the body. Here, the heart's ejection fraction (EF) number[4], a measurement of its pumping capacity, is reduced. The second case of left-sided heart failure is known as diastolic failure. This occurs when the left ventricle loses its ability to relax normally, often due to muscle stiffness. As a result, it is unable to fill with enough blood during the heart's resting period and the supply of oxygen-rich blood is limited again. People with diastolic heart failure can have a normal ejection fraction number, which makes them more complicated to diagnose and more difficult to treat [21] [22].

The second type of heart failure, right ventricular heart failure in this case, takes place when the right side of the heart starts to lose its ability to pump used blood back to the lungs for oxygen replenishment. When the right side loses its pumping ability, blood backs up in the body's veins. This can then cause swelling or congestion in the legs or ankles and swelling within the abdomen [21]. Right-sided heart failure is often caused by an initial left-sided failure. When

[4] Ejection fraction is a measurement of the percentage of blood leaving the heart each time it contracts with a heartbeat. According to the Mayo Clinic, an LV ejection fraction of 55% or higher is considered normal. An LV ejection fraction of 50% or lower is considered reduced. Experts vary in their opinions about an ejection fraction between 50% and 55%.

the left ventricle fails, increased fluid pressure is transferred back through the lungs, which damages the heart's right side.

Treatments are different for different types of heart failure, and the medical community has come a long way in the understanding, diagnosis, treatment and prevention of heart failure acute episodes. In the last three decades, there have been major advances in heart diagnostic tools like sophisticated echocardiography and cardiac magnetic resonance imaging. In addition, better blood markers exist today to diagnose and monitor patients with heart failure. For example, the B-type natriuretic peptide, or BNP, is a protein secreted by the heart's ventricles when the heart failure condition worsens. As a result, BNP blood tests can also be used to differentiate shortness of breath caused by heart failure from that caused by pneumonia. In addition, drugs have expanded and now include ACE (angiotensin-converting enzyme) inhibitors for heart failure and beta-blockers. ACE inhibitors expand blood vessels and allow the heart to function more efficiently. Beta-blockers reduce the heart's workload and, as a result, help with managing heart failure [22].

Moreover, innovative devices such as defibrillators and biventricular pacemakers exist today and are capable of monitoring heart rhythm. By sending electrical impulses, implanted biventricular pacemakers are able to synchronize the contractions of the right and left ventricles of the heart and, thus, normalize heart rhythm. Similarly, implantable cardiac defibrillators are implanted for activation during arrhythmia or tachycardia emergencies, where an electrical charge is delivered to the heart to return its rhythm to normal. Incorporating such devices in the monitoring of heart failure patients, where possible, can have a significant effect on improving heart failure patient outcomes [22] [23].

2.2 Strategies for Enhanced Heart Failure Patient Care

A broad review of the existing literature was also carried out to understand the different strategies that hospitals and health systems use to provide enhanced care to their heart failure patients. In particular, an emphasis was placed on identifying common themes related to outpatient monitoring and heart failure transition-of-care models that can minimize exacerbation and hospitalizations. Six common themes were identified [24].

Theme 1: Planning for Discharge and Outpatient Follow-ups [25] [26] [27]

This theme is focused on extending the time period between hospital admission episodes. It revolves around the importance of starting the discharge planning process from the day of hospitalization. The process includes measuring clinical indicators during the hospitalization and assessing physical signs and symptoms to ensure adequate decongestion prior to discharge. In addition, assessing personal, social, economic and cultural factors are also important as they are all expected to have an effect on the patient's lifestyle and the time of their next hospitalization. These include key factors like patient's lifestyle, tendency to adhere to drug therapy, and general awareness of evolving symptoms and disease progression.

Moreover, post-discharge outpatient follow-ups also fall under this theme. A common strategy employed by hospitals to manage outpatient follow-ups is the establishment of specialized nurse-led heart failure clinics. In these clinics, diuretics and other heart failure medications may be administered as needed. In addition, these clinics also serve the need of scheduling post-discharge follow-ups within 7-14 days of discharge. During these follow-ups, patient education is reinforced, barriers to social and community supports are assessed, and advanced-care planning is discussed. As part of hospital strategies to improve heart failure

transition-of-care, discussions between health care team members and patients about the rationale for early (within 7–14 days) follow-ups with cardiologists or specialized heart failure clinics are embedded in the planning for discharge process.

Theme 2: Multidisciplinary Teamwork, Communication and Coordination [28] [29] [30] [31] [32] [33] [34]

The importance of team communication is stressed and this includes communication between the multidisciplinary health care providers (including patients' primary care providers and cardiologists), patients, and family members (or other caregivers). Communication failures, which are associated with delays in diagnosis and treatment, are to be avoided through complete and standardized information transmission. Checklists and electronic records are presented as useful tools to aid with standardized information transfer.

Theme 3: Organized Information Collection and Medication Reconciliation [35]

Collecting patient information in a timely and organized manner was another strategy followed by hospitals for enhanced heart failure patient care. Components of these transition or discharge records typically include: reason for hospitalization; procedures performed during hospital stay; treatments and/or services provided during hospital stay; discharge medications; care team members involved and their information; and follow-up treatments or services.

Medication reconciliation is particularly important within the context of information collection and organization. Here, the importance of medication monitoring and tracking is emphasized to identify and prevent any possible drug interactions with a potential for adverse outcomes.

Theme 4: Engaging Social and Community Support Groups [36]

The fourth theme identified in the literature is related to the importance of engaging with social community services. Such programs often complement traditional medical care by providing heart failure patients with assistance in household activities, meals and other necessities. In fact, one study reported that social support was associated with increased adherence to medications, diet and other care aspects related to a patient's lifestyle.

Due to their benefits, hospitals are encouraged to utilize support from social and community services. For an effective engagement, health care providers are encouraged to take the time to understand patient needs and how they may interfere with their self-care choices. This would then help hospitals assess potential barriers that heart failure patients face regarding the utilization of social and community services, and their reluctance to accepting them.

Theme 5: Patient Monitoring and Education [37] [38] [39] [40]

As a result of the chronic nature of heart failure and its heavy dependence on lifestyle factors, this theme was one of the most significant ones in improving long-term outcomes for heart failure patients. Monitoring for new or worsening heart failure signs and symptoms can lead to early detection of worsening patient states, intervention related to self-care management and effective treatment. This would then reduce all-cause admissions, mortality rates and outpatient or emergency care visits. Due to the complexity of monitoring and its significance on patient indicators, this theme motivated the analyses and models presented in later sections. Moreover, in Section 2.3, we explore the area of remote monitoring in greater depth and focus on relevant clinical signs for monitoring as well as the systems, processes and workforces supporting and acting on the collected information.

Theme 6: Advanced Care Planning, Palliative and End-of-Life Care [41] [42]

While not directly related to this project, the importance of advanced care planning discussions, to correct any misperceptions about the long-term and progressive nature of heart failure, served as the sixth theme related to hospital strategies around improving heart failure outcomes. These discussions help in highlighting the important role that advanced therapies play by reinforcing the importance of adhering to self-care recommendations that stabilize heart failure. As such, setting realistic expectations around patient prognosis and discussing advanced-care planning regularly during the course of care – and not just at the point of transition to mechanical circulatory support, transplantation, or palliative care – is essential.

2.3 Remote Monitoring for Heart Failure

As discussed in Theme 5 of Section 2.2, close monitoring of heart failure signs can help with the improvement of patient outcomes. In this section, we start with a generic overview of the feedback loops involved in home heart failure monitoring and management. This then motivates the aspects through which we discuss and compare three different remote monitoring schemes.

Remote monitoring is defined as "a strategy that provides the patient with some sort of education around self-care, or monitors patients via different devices for early detection of heart failure decompensation and intervention" [43]. This can take place in three ways: (i) structured telephone support, where contacts between patients and healthcare providers, that may or may not include transfer of physiological data, take place; (ii) telemonitoring, where transfer of patient physiological data (e.g., blood pressure, weight, electrocardiographic details, and oxygen saturation levels) takes place through telephone or digital transfer from a patient's home to a healthcare provider; or (iii) a combination of structured telephone support and telemonitoring.

Although not available for all patients and not fully integrated with the MGH cardiology department, current remote monitoring, or telemonitoring, techniques at Partners HealthCare[5] are presented as the first monitoring scheme in Section 2.3.2. Next, in Section 2.3.3, we present two alternative and more advanced monitoring techniques that are obtained from a broad literature review. These include: (i) monitoring through implantable hemodynamic devices and (ii) monitoring through daily concentrations of B-type natriuretic peptide (BNP).

2.3.1 Generic Overview of Home Heart Failure Management

In order for a monitoring strategy to be effective, it needs to be looking at the 'right' physiological parameters that could (i) allow the care team to predict heart failure decompensation events accurately and finely and (ii) permit timely intervention. In addition to monitoring clinically relevant indicators, effective monitoring strategies also require efficient processes for data transmission and interpretation. As such, the systems and workforce supporting and acting on monitored information represent another key factor in determining the success of home heart failure monitoring systems. A generic representation of the closed-loop feedback system in home heart failure management systems is presented in Figure 2-2 [44]. Each entry in this figure represents an area through which any monitoring strategy could be assessed.

The monitoring care path starts with the measurement of some physiologic indicator(s) that has/have been identified as reliable signals for the early detection of heart failure deterioration. This data is then transmitted to the patient and/or relevant care provider (e.g.,

[5] Brigham and Women's Hospital and Massachusetts General Hospital founded Partners HealthCare in 1994. Today, Partners is an expanded network of health centers and providers in Massachusetts. In addition to its founding hospitals, some of the other hospitals it covers include McLean Hospital, Newton-Wellesley Hospital and North Shore Medical Center [62].

primary care physician, cardiologist, nurse practitioner or physician's assistant). Typically, the care team member would then interpret the data and advise on a corrective intervention and therapeutic recommendation where necessary. Upon implementing the recommended change, the patient is then reassessed for response. To determine if a desired change in status has taken place, the same physiologic indicator(s) is/are measured again and the loop repeats itself.

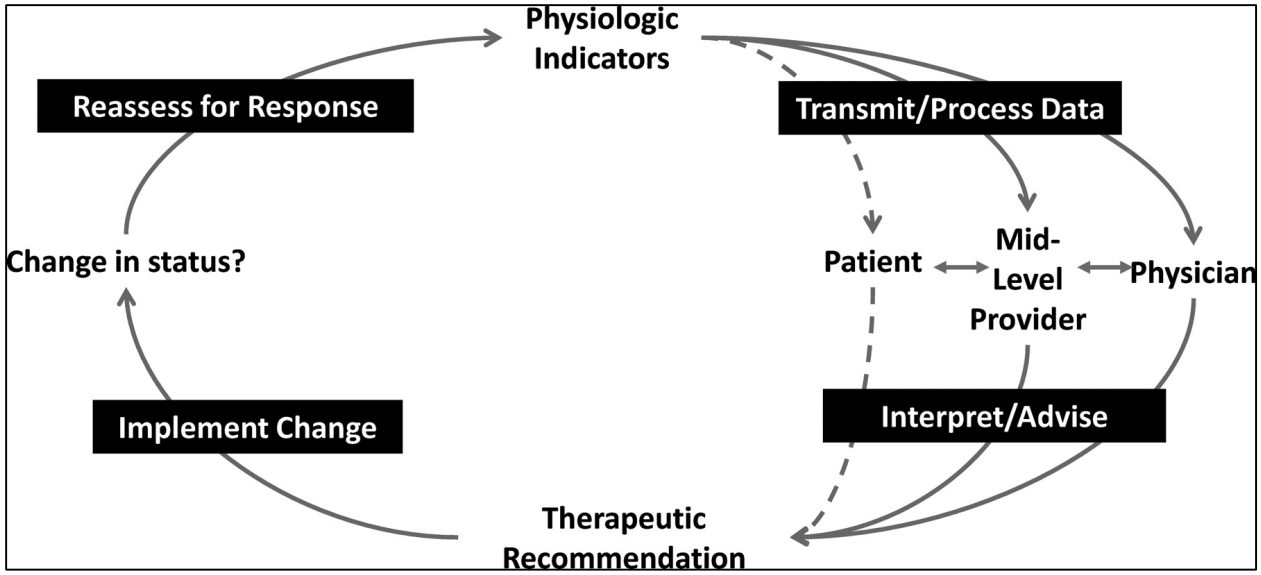


Figure 2-2: Generic Representation of a Closed-Loop Feedback System in Home Heart Failure Management

We now consider three forms of remote monitoring that are based on different physiologic indicators. For each, we primarily discuss: (i) the signs and symptoms being monitored and their clinical relevance, and (ii) the systems and workforces supporting and acting on monitored information. For all remote monitoring techniques, the primary corrective intervention available to health care providers, besides reinforcement of education around lifestyle choices and self-management, is an adjustment of diuretic therapy.

2.3.2 Telemonitoring at MGH: The Connected Cardiac Care Program

At MGH, telemonitoring is available under the Partners HealthCare at Home Program, and is known as the *Connected Cardiac Care Program* (CCCP). It was designed as a home monitoring and education program that aims to improve self-management of heart failure patients within the Partners network of hospitals that have a risk of hospitalization. Enrollment in CCCP is selective and takes place for four months (renewal is required for longer time periods). To be enrolled, a patient has to meet the following list of eligibility criteria [43]:

1. Patient is enrolled in the Integrated Care Management Program (iCMP)[6]
2. Patient resides in a Partners HealthCare at Home service area
3. Patient's provider at MGH and case manager are both in agreement with the service
4. Patient has Class II, III, or IV heart failure (refer to footnote in Section 1.2.1 for details)
5. Patient has a moderate or high risk for hospitalization (i.e., iCMP risk score of 1 or 2)
6. Patient is able to communicate in either English or Spanish
7. Patient either has a competent and willing caregiver or can assume full personal responsibility of telemonitoring
8. Patient's home setting offers a clean, safe environment for the equipment and either a phone line or cell modem

In 2015, a study[7] was carried out on 348 MGH patients to understand the approach of CCCP and evaluate its effect on hospitalization and mortality [43]. Each day, the participants

[6] The Integrated Care Management Program (iCMP) is a Partners HealthCare program that matches high-risk and chronically ill patients with nurse care managers that closely monitor them during and after office appointments using phone calls and home visits [63]. The iCMP scores represent a general risk score of hospitalization and are independent of a heart failure condition.

[7] The study involved a retrospective database review of medical records of patients.

monitored the following four physiologic parameters (as expected of CCCP enrollment): (i) blood pressure, (ii) heart rate, (iii) weight, and (iv) blood oxygen saturation. They then answered questions on heart failure symptoms on a touch-screen computer. Monitoring equipment, as in typical CCCP enrollment, included ViTel Net and devices approved by the FDA: a UA 767PC Turtle 400 monitor, a Life-Source digital weight scale, an A&D blood pressure cuff and meter, and a BCI pulse oximeter device (UC-321PBT).

In addition to monitoring physiologic parameters and answering daily questions about symptoms, participants also received structured biweekly telephone-based education sessions over their 8-week enrollment period. As an outcome, the study reported an association between CCCP and significantly lower hospitalization rates up to 90 days and significantly lower mortality rates over 120 days of the program. The effects, however, did not persist beyond the 120-day program duration. This suggested two things: either (i) participants acquired a dependency on CCCP, or (ii) patients did not have enough time to develop sufficient self-competency for disease management over the assigned enrollment period of only four months.

Figure 2-3 depicts the CCCP process, where home monitoring is initiated by the patient reporting physiologic and other data. This is then followed by data collection in a data repository platform within Partners, patient data evaluation by a qualified nurse, and, finally, care coordination with a medical doctor as needed.

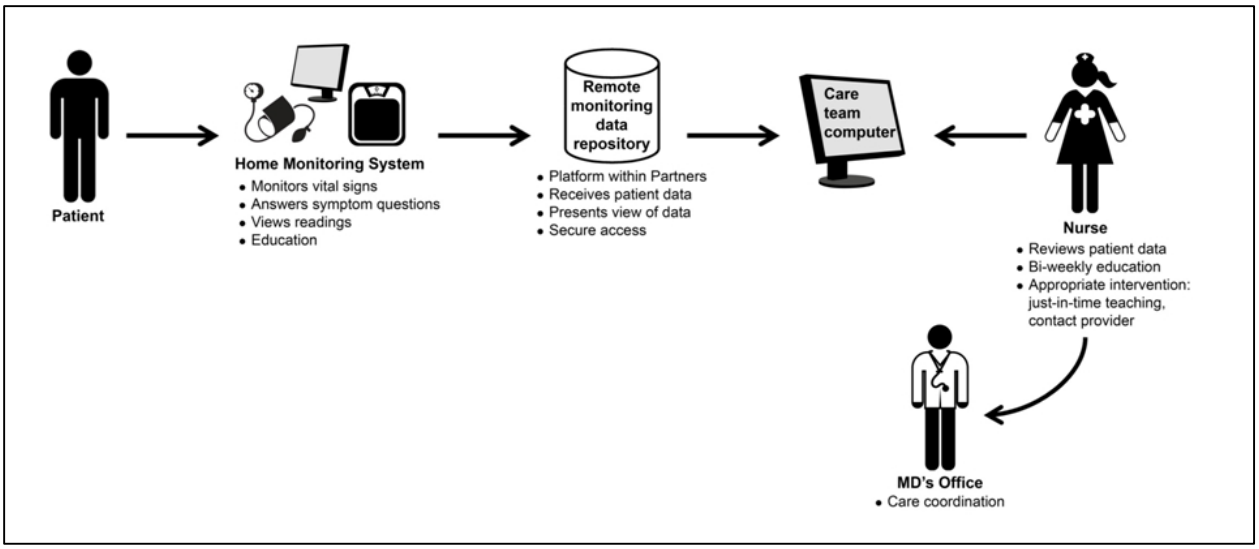


Figure 2-3: Connected Cardiac Care Program *[43]*

2.3.3 Remote Monitoring through Implantable Hemodynamic Monitors

Although current medical practices around heart failure management are far from being automated, the future of heart failure management may actually reside in monitoring through implantable devices, potentially in combination with other intracardiac devices. These devices can provide continuous measurements of intracardiac filling pressures[8], which have been validated as reliable signals for the early detection of heart failure deterioration [45].

In fact, clinical trials have already been performed in this area. Perhaps the most famous study is the CHAMPION trial, which was designed as a randomized clinical trial to test the

[8] Intracardiac filling pressures are mainly characterized by the heart's 'right-side filling pressure' and 'left-side filling pressure'. The former represents pressure in the heart's right atrium during contractions while the latter represents pressure in the left atrium. These pressures fill the ventricles during the heart's resting phase (diastole) between contractions. Intracardiac filling pressures are mainly influenced by 'intravascular volume' where, during dehydration, filling pressures decrease and, during fluid overload, filling pressures increase. Other factors like disorders of ventricles and their valves can also influence intracardiac filling pressures.

hypothesis that heart failure management, based on frequently measured pulmonary artery pressures, is superior to traditional monitoring methods (i.e., weight-based monitoring techniques) [46]. In the trial, a total of 550 subjects, with Class III heart failure, received the CardioMEMS implant, an FDA-approved wireless sensor that monitors filling pressures. Patients were then randomized to treatment with or without the assistance of sensor data. After six months, the results showed that care management guided by the pressure sensor was associated with a statistically significant 30% reduction in heart failure hospitalizations. These results were attributed to the increase in the number of adjustments to heart failure medications by physicians who had access to the sensor data (2,468 adjustments; mean 9.1 adjustments per patient) relative to those without access to the data (1,061 adjustments; 3.8 adjustments per patient). As such, the results suggested that, for patients with advanced heart failure symptoms and implanted hemodynamic monitors/sensors, continuous monitoring of intracardiac pressures may provide a reliable signal for early detection of heart failure deterioration and early intervention [44]. The main challenge here is that such an approach requires invasive surgeries to insert the sensor implants, which might not appeal to all patients. This is further complicated by the high costs of such procedures and the required infrastructure development for effective implementation (i.e., "who monitors the monitor").

2.3.4 Remote Monitoring through Daily Concentrations of B-type Natriuretic Peptide (BNP)

The HABIT clinical trial is another study presented in the literature with the purpose of determining more accurate clinical predictors of acute heart failure decompensation (ADHF). Specifically, the aim was to determine how BNP concentrations[9] correlate with ADHF [47].

The study was performed on a total of 163 patients with heart failure signs and symptoms of ADHF. The patients, who represented a mix of hospital discharged cases and outpatient cases, measured their weight and BNP levels daily for 60 days with a finger-stick test. Both patients and health care providers were blinded to BNP levels during this period, so as to avoid corrective actions in response to this specific predictor. The measured outcomes were the following decompensation events: (i) cardiovascular death; (ii) hospital admission because of decompensated heart failure; or (iii) clinical heart failure decompensation requiring changes in heart failure related therapy. The final results indicated that there exists a positive correlation between BNP and decompensation risk. Furthermore, the study advised that home BNP testing is feasible and that it is most promising in monitoring applications when used as a complement to daily weight measurements [47].

Finally, before being approved as an acceptable way of monitoring heart failure patients remotely, the findings from the HABIT clinical trial would need to be verified on larger groups of patients through additional trials and studies. Should it prove successful, perhaps it would appeal to patients that do not desire the invasive surgeries that hemodynamic remote monitoring

[9] B-type natriuretic peptide (BNP) is a neurohormone secreted mainly in the cardiac ventricles in response to volume expansion and pressure overload. Data have shown that BNP levels correlate with the severity and prognosis of heart failure [64].

requires. In terms of systems and workforce supporting and acting on monitored information, we expect BNP-based remote monitoring to be of a similar nature to that of CCCP, perhaps even complementing CCCP's monitored signs with an additional clinically-relevant parameter.

2.4 Predictive Modeling of Hospitalization Risk

Remote monitoring was presented in Section 2.3 as a potential way of detecting decompensation signs, based on a 'real-time' monitoring of various clinical parameters. However, due to the high costs and technical challenges associated with implementing remote monitoring on a large scale, hospitals and other care systems have also considered alternative methods of identifying patients at risk of decompensation and hospitalization. One such alternative is the use of machine learning techniques to predict hospitalization risk, based on patients' historical records and billing data.

In this section, we present an overview of some heart failure related (and all-cause) predictive risk models that exist in the literature[10]. We base our discussion on the following three factors: (i) the outcome that each model aims to predict; (ii) the data and predictive features used in each model; and (iii) the reported performance of each model.

2.4.1 Examples of Common Heart Failure Risk Models

In an article published in *Medical Care* [48], Greenwald et al. present a novel model that predicts 30-day all-cause rehospitalization risk. In addition to clinical variables, they include factors describing physical function, cognitive status and psychosocial support. To evaluate the

[10] The predictive models are also presented as a point of contrast with our first-time heart failure admission model. We specifically present these models for discussion for one of the following two reasons: (i) they are widely used and accepted as predictors of heart failure risk; or (ii) they guided the development of our model and its features.

performance of their final model, which consists of a total of 16 variables, they use out-of-sample testing and report an area under the receiver operating curve (AUC) ranging between 0.70 and 0.75 (refer to Section 0 more details on AUC as a performance metric).

In another report [49], Cronin et al. discuss the development and implementation of a real-time 30-day all-cause readmission predictive model at MGH. Their model, which was developed using MGH admission data over the course of two years, includes a total of nine predictors and has a reported real-time implementation AUC[11] of 0.671.

Finally, we note the Seattle Heart Failure Model (SHFM), designed to predict 1- to 3-year survival rates. SHFM is perhaps one of the most widely used models in predicting heart failure risk. Developed using a cohort of 1,125 heart failure patients, it uses clinical, pharmacological, device and laboratory parameters as predictive features [50]. Its overall reported performance is an AUC of 0.729. Later, in Section 5.4.5, we compare the SHFM against our final model in greater depth.

2.4.2 Addressing Limitations in Existing Risk Models

By somehow identifying 'high-risk' patients, all of the aforementioned models essentially provide some method of prioritizing patients in greatest need for the limited resource of proactive outpatient provider care, attention and monitoring. However, the first two models are designed to predict 30-day readmission as an outcome and, hence, fail to predict *de novo* or "first-time" admissions by definition. On the other hand, while the third model is not necessarily limited to readmissions, it predicts risk on the different outcome of survival rate. Our model

[11] The model was developed using retrospective data from 45,924 MGH admissions between 2/1/2012 and 1/31/2013. It was then validated prospectively in a real-time implementation for 3,074 MGH admissions between 10/1/2013 and 10/31/2013.

takes care of these issues by using some of the features that these models employ (in addition to other new ones) and predicting heart failure risk on the basis of probability of occurrence of a first-time admission in a given year.

The results and details of our model are discussed in greater depth in Chapter 5. As is common in other problems tackled through a predictive analytics approach, a four-step process was taken to arrive at the results. These include: (i) identify risk factors; (ii) train a model and predict first-time admissions; (iii) validate the model using out-of-sample testing; (iv) define interventions using the proposed model.

To identify some risk factors as potential predictive features (in addition to the ones already obtained from previous work efforts), we carried out an exhaustive mapping of the heart failure care pathway that aimed to generate an understanding of the flow of heart failure patients and their interactions with the system. This is discussed next in detail in Chapter 3.

3. Heart Failure Care Management at MGH

3.1 High-Level Current-State Care Pathway

This chapter is based on a process mapping of the current care pathways of heart failure patients. We start with a high-level mapping, shown in Figure 3-1, which consists of four main sections: (i) outpatient identification and care; (ii) emergency department care; (iii) inpatient care; and (iv) post-discharge outpatient care, monitoring and escalation.

In Section 3.2, we present an overview of the data sources accessed and used for our analyses, before proceeding to a more in-depth discussion of each of the four aforementioned areas of the heart failure care pathway (in sections 3.4 to 3.6). Finally, in Section 3.7, we conclude by presenting the key hypotheses that were generated during the mapping activities and motivated the quantitative analyses discussed in Chapter 0.

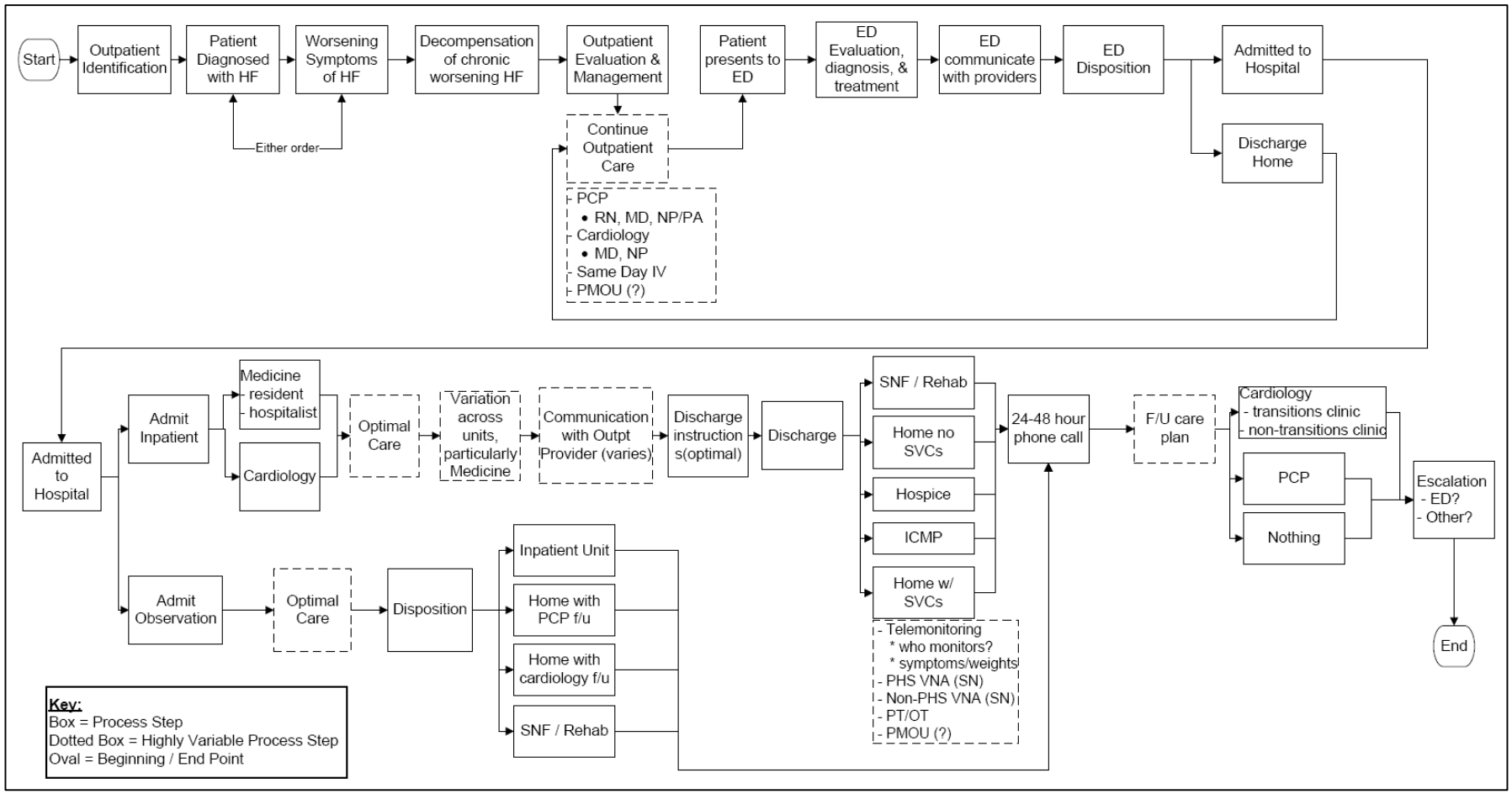


Figure 3-1: High-Level Heart Failure Care Pathway at MGH

3.2 Data Sources

Several hospital data sources were accessed and used for the purposes of this project. The hospital billing system, EPSi, and the Massachusetts General Physicians Organization (MGPO) billing system were used to identify the primary diagnoses associated with both hospital and outpatient encounters, as described in the following section. For inpatient encounters, historical data on each hospitalization was obtained from EPIC, specifically from the 'readmissions table'. Similarly, EDIS was used to identify ED-specific encounters and their relevant details. The general time frame followed was January 2011 to September 2015.

To analyze institutional ownership in terms of 'PCP ownership', we used the MGH Laboratory of Computer Science (LCS) Dynamic Linkage Cohorts (DLC) system to identify the MGH patients that are actively linked to MGH PCPs. For institutional ownership in terms of 'cardiologist ownership', we considered outpatient encounters with MGH cardiologists as they appeared in MGPO billing.

For analysis of outpatient appointments (Chapter 0), Cadence Scheduling was used between October 2014 and September 2015. Prior to this time frame (specifically, from 2009 to July 2014), the appointment record system in use for outpatient clinics was IDX. Both Cadence and IDX contain a record for all scheduled outpatient appointments in different clinics, including those that were completed, canceled and re-scheduled. However, due to the differences in recording styles and the desire to explore more recent trends in outpatient appointments, IDX appointments were not analyzed as part of this project.

For the predictive model (Chapter 5), the Research Patient Data Registry (RPDR) was also consulted for data on patient demographics, medications, health histories and relevant providers. With regards to composition and structure, The RPDR Database is composed of over

6.5 million patients and over 2 billion records from patient encounters, labs and results, and other medical care. This includes services that patients received at either MGH or other hospitals within the Partners network. As a first iteration of this study, we focus only on MGH services and encounters. Finally, for one of the model's predictors, specifically 'median income by zip code', we used income data published on the U.S. Census Bureau's official website.

3.3 Outpatient Identification and Care

In the outpatient setting, the heart failure care path usually starts with the identification and diagnosis process[12], as shown in Figure 3-2. When symptoms arise, a patient typically reaches out to his/her primary care provider, often through a phone call. Following some form of assessment, the provider would typically then decide whether or not the patient can adjust to therapy (e.g., increase medication or diuretic dosage) and would confirm with the patient whether or not that is needed.

In the event that a provider decides that the patient cannot adjust to therapy, then he/she is followed up with in one of several ways: (i) schedule a same-day appointment; (ii) schedule an appointment within a few days (and manage by phone until then); (iii) refer to a cardiologist (if not seeing one already); or (iv) send the patient to the ED. As shall be discussed subsequently in Chapter 0, current-state data analyses show that outpatient clinics, particularly within cardiology, often face issues around limited access, which results in longer-than-desired wait times for patient appointments.

[12] When patients present with heart failure symptoms, they are classified into one of two diagnosis groups: (i) newly identified with a heart failure diagnosis (i.e., no heart failure diagnosis code appeared in previous encounters); or (ii) classified with pre-existing and, potentially, decompensating heart failure (i.e., a heart failure diagnosis code had appeared in at least one previous encounter).

Before delving into the details of limited outpatient access, we proceed with a classification of heart failure patients, based on their 'institutional ownership', and a continuation of the process mapping findings in other areas of the heart failure care pathway.

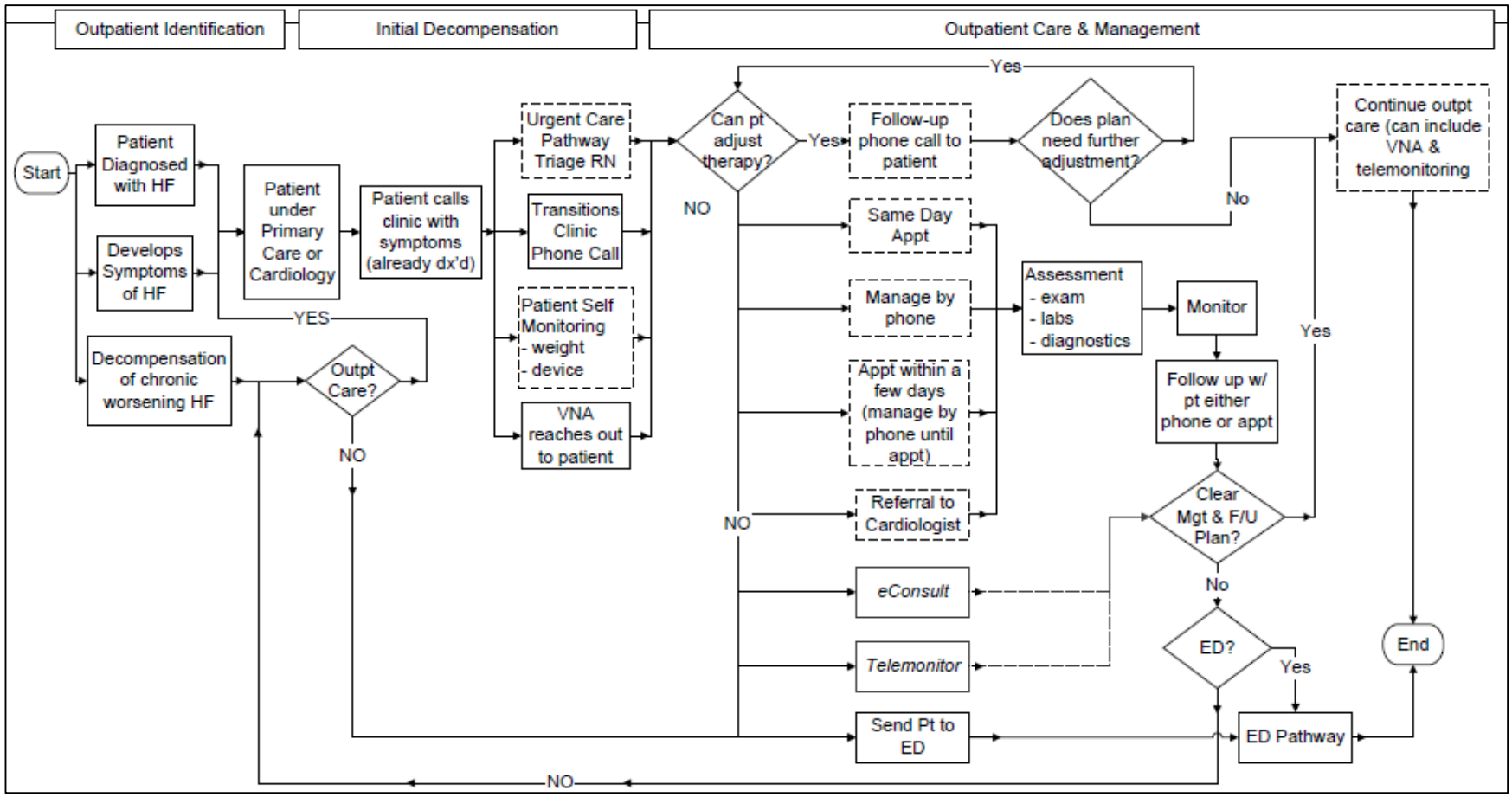


Figure 3-2: Pre-Admission Process: Outpatient Identification, Care and Management

3.3.1 Classification of Heart Failure Patients

While mapping the current-state care pathway of heart failure patients, the team realized that there was no form of heart failure registry capturing all patients that appeared with a heart failure diagnosis code in the inpatient and/or outpatient setting of MGH. As such, the first priority in terms of data analysis was to generate a comprehensive list of heart failure patients that had inpatient and/or outpatient encounters at MGH[13].

To limit encounters to heart failure ones, we turn to the International Classification of Diseases (ICD) codes in their ninth and tenth versions (i.e., ICD-9 and ICD-10) and derive primary heart failure diagnosis codes[14] from EPSi and MGPO billing systems. Between October 2012 and September 2015, we find a total of 11,301 unique patients with at least one encounter that is associated with a heart failure primary discharge code (in the inpatient and/or outpatient setting of MGH). Table 3-1 shows the total number of heart failure patients appearing in each of fiscal year (FY) 2013, 2014 and 2015. Each fiscal year starts with the October of the previous year and ends in September of the year of interest. Patients are grouped by their encounters in the inpatient (IP) setting only, outpatient (OP) setting only, or both (IP and OP).

Furthermore, to distinguish between one-off visits and patients that are 'within the MGH network', institutional ownership is defined using patient engagement with MGH-affiliated primary care providers (PCPs) and cardiologists (as explained earlier in Section 3.2). In Table

[13] It is important to highlight that all analyses presented in this thesis are limited to MGH data only. Activities in other settings (including the Partners HealthCare) are unknown at this point. With the availability of more data from the broader Partners network and beyond, this analysis can be extended to account for patient interactions beyond the MGH network.

[14] Heart failure diagnosis codes were already mapped out as part of an earlier MGH project. The same ICD-9 and ICD-10 codes were used for our project and are listed in Appendix B.

3-1, 'PCP-owned' patients (as determined by the LCS DLC algorithm) appear in Groups 1 and 2. Groups 2 and 3 consider patients 'owned by MGH cardiologists', where ownership here is defined as having had an outpatient evaluation and management encounter with an MGH-affiliated cardiologist in the 12 months prior to the year of interest. Finally, Group 4 in Table *3-1* represents those patients that have no form of institutional ownership.

The overall goal of this exercise is to develop a method of classifying heart failure patients[15] based on their care patterns. More specifically, we consider two factors: (i) institutional ownership (i.e., 'who cares for and interacts with them regarding the heart failure condition'); and (ii) the patients' outpatient and/or inpatient encounters.

Table 3-1: Identifying Heart Failure Patients at MGH: (i) FY2013 (Oct '12 – Sep '13); (ii) FY2014 (Oct '13 – Sep '14); (iii) FY2015 (Oct '14 – Sep '15)

Group 1: Owned by MGH PCPs only **1,093 patients (20.9%)**		
IP Only	IP and OP	OP Only
88 patients (1.7%)	201 patients (3.8%)	804 patients (15.4%)

Group 3: Owned by MGH Cardiologists only **1,592 patients (30.5%)**		
IP Only	IP and OP	OP Only
0 patients (0%)	148 patients (2.8%)	1,444 patients (27.6%)

Group 2: Owned by MGH PCPs & Cardiologists **1,167 patients (22.3%)**		
IP Only	IP and OP	OP Only
0 patients (0%)	188 patients (3.6%)	979 patients (18.7%)

Group 4: Not Active Within MGH Network **1,347 patients (26.3%)**		
IP Only	IP and OP	OP Only
184 patients (3.5%)	166 patients (3.2%)	1,024 patients (19.6%)

(i)

[15] In addition to direct impact on the work and analyses described in this thesis, the greater value of a classification system lies in its ability to serve as a foundation for the development of a heart failure patient registry that captures each patient's "owner(s)" and facilitates communication between patients and healthcare providers/specialists.

Group 1: Owned by MGH PCPs only **1,154 patients (19.4%)**		
IP Only	IP and OP	OP Only
94 patients (1.6%)	207 patients (3.5%)	853 patients (14.3%)

Group 3: Owned by MGH Cardiologists only **1,984 patients (33.4%)**		
IP Only	IP and OP	OP Only
0 patients (0%)	187 patients (3.1%)	1,797 patients (30.2%)

Group 2: Owned by MGH PCPs & Cardiologists **1,456 patients (24.5%)**		
IP Only	IP and OP	OP Only
0 patients (0%)	205 patients (3.4%)	1,251 patients (21.0%)

Group 4: Not Active Within MGH Network **1,352 patients (22.7%)**		
IP Only	IP and OP	OP Only
175 patients (2.9%)	223 patients (3.8%)	954 patients (16.0%)

(ii)

Group 1: Owned by MGH PCPs only **1,228 patients (20.4%)**		
IP Only	IP and OP	OP Only
126 patients (2.1%)	203 patients (3.4%)	899 patients (14.9%)

Group 3: Owned by MGH Cardiologists only **2,054 patients (34.1%)**		
IP Only	IP and OP	OP Only
0 patients (0%)	186 patients (3.1%)	1,868 patients (31.1%)

Group 2: Owned by MGH PCPs & Cardiologists **1,577 patients (26.2%)**		
IP Only	IP and OP	OP Only
0 patients (0%)	197 patients (3.3%)	1,380 patients (22.9%)

Group 4: Not Active Within MGH Network **1,156 patients (19.2%)**		
IP Only	IP and OP	OP Only
264 patients (4.4%)	137 patients (2.3%)	755 patients (12.6%)

Key:

Ownership by MGH PCPs

Ownership by MGH Cardiologists

(iii)

In addition to identifying heart failure patients, we also consider the dynamics of heart failure patient appearances in inpatient and outpatient billing data. This acts as the first step towards

understanding the frequency of engagement that patients are likely to have with their providers. A schematic of the profile of heart failure diagnosis appearances for patients affiliated with MGH PCPs is shown in Figure 3-3. Clearly, the system is highly dynamic with most patients appearing as newly diagnosed and a large number of patients not appearing every year. This dynamic nature in patient classification is due to possible changes in clinical conditions (e.g., hospitalization or death) and/or changes in institutional ownership (network association).

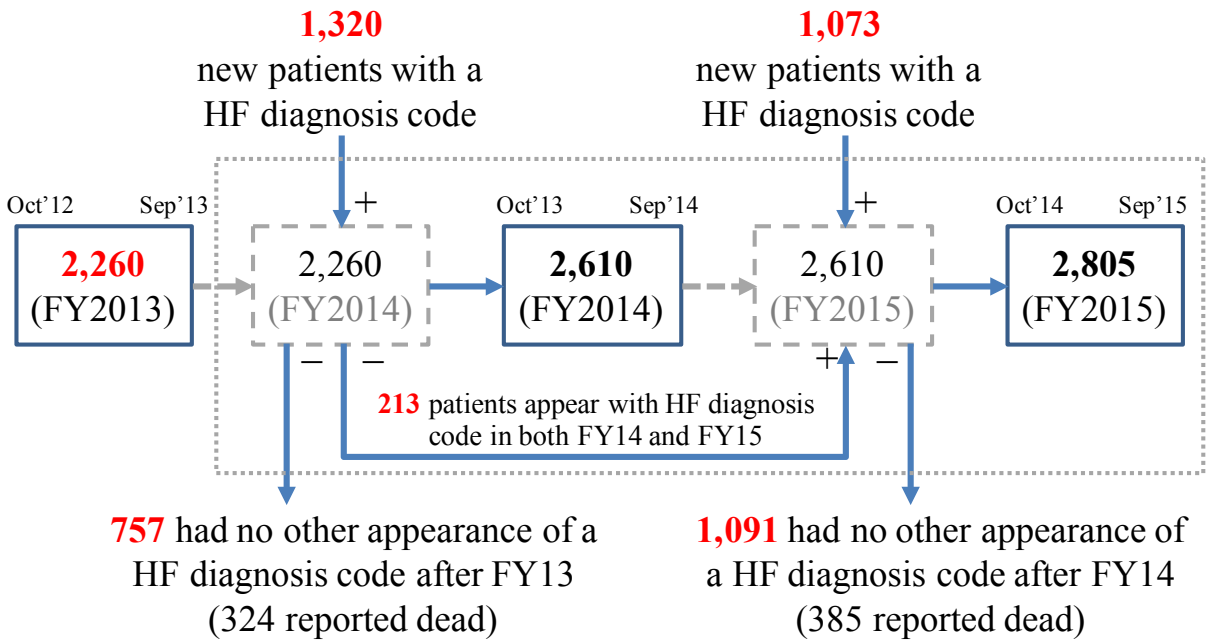


Figure 3-3: Dynamics of Heart Failure Diagnosis Code Appearances for Patients Affiliated with MGH PCPs; Time Frame: Oct '12 – Sep '15

3.4 Emergency Department

3.4.1 Emergency Department Current-State Care Pathway

The ED serves as the greatest source of heart failure hospital admissions. Admissions to the ED, shown in Figure 3-4, can happen in several ways: (i) referral from a primary care physician

(PCP) or specialist; (ii) transfer from another clinic/facility; or (iii) patient self-presentation (walk-in).

Once a patient presents to the ED, he/she is triaged and a note is generated. The patient would either be classified as: (i) patient with pre-existing heart failure; or (ii) patient with newly diagnosed heart failure. In addition, each patient is coded with either 'mild congestive heart failure (CHF) and shortness of breath (SOB)' or 'acute decompensating heart failure with SOB'. Next, the patient is seen and assessed by an ED care team member who may or may not communicate with the referring providers (if any).

Based on the ED provider's assessment, the patient can take one of several pathways: (i) admitted to the hospital with either the general medicine service or the cardiology service; (ii) transferred to the ED observation unit; (iii) treated in the ED and discharged with home outpatient care (rare); or (iv) treated in the ED and discharged with home services (e.g., Visiting Nurse Association (VNA); Partners Mobile Observation Unit (PMOU); or rehabilitation services[16]). While the various ED discharge dispositions were not analyzed in great depth under this project's scope, understanding that patients can take various paths following their discharge is important. In particular, patients receive variable care based on their various possible discharge dispositions and that is a source of variation that can potentially affect when a discharged patient will have their next ED visit or hospital admission.

[16] Refer to Appendix A for a brief description of each of these types of home services.

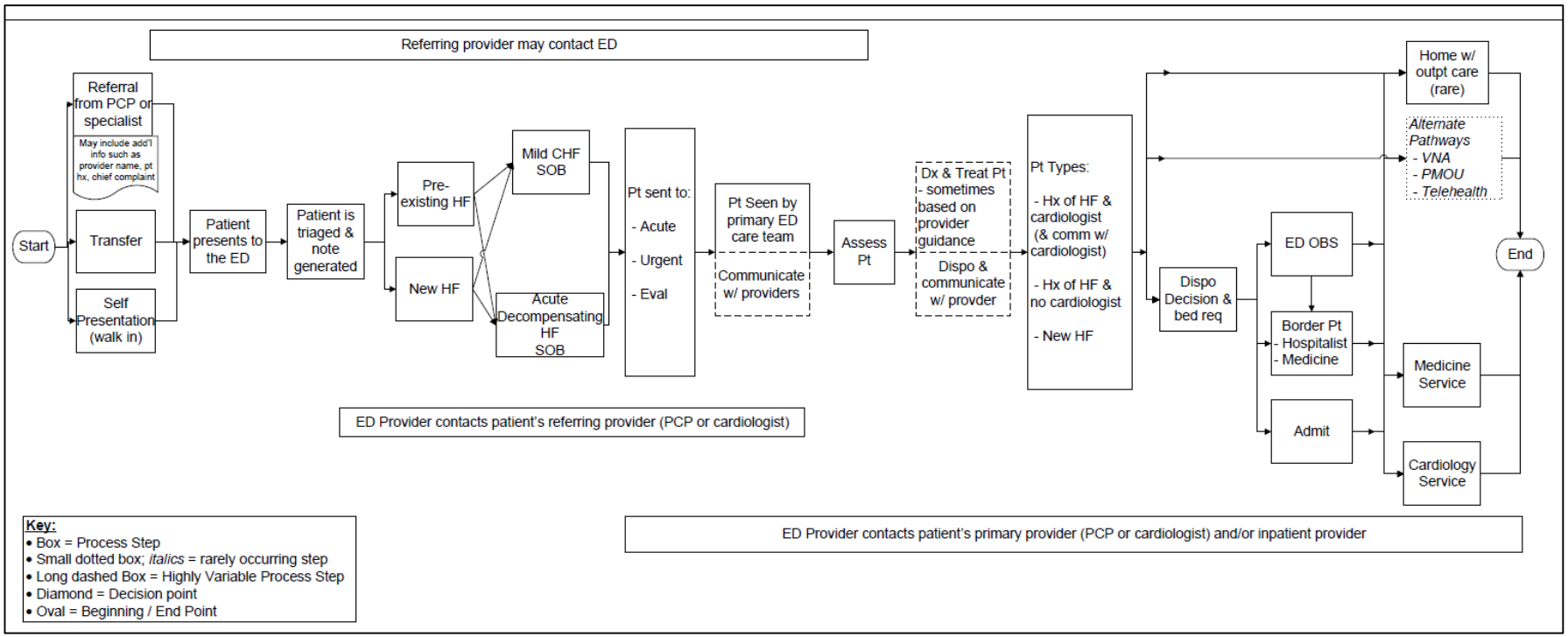


Figure 3-4: Emergency Department Care Pathway

3.4.2 Distribution of Sources of Admission

To confirm whether or not "most heart failure admissions originate at the ED", we consider ED encounter data and hospital admission data from EDIS and EPSi, respectively, between October 2012 and September 2015 (i.e., FY2013 to FY2015). From EPSi, we extract 3,801 patient encounters[17] that were coded with a heart failure primary discharge diagnosis at least once during this time frame. We then create a merge (or inner join) between the EDIS and EPSi datasets on the following two conditions: (i) matching patient medical record numbers (MRNs); and (ii) hospital admission date (from EPSi) equals ED discharge date (from EDIS)[18]. Upon merging the datasets, the resulting number of data rows drops from 3,801 to 2,824, meaning that 2,824 (or ~74.3%) of the admission encounters also appeared as ED encounters on the same day. We therefore confirm the hypothesis that the *majority of hospital admissions originate at the ED*, and turn to analyzing the sources of ED referrals leading to hospital admissions next.

To get a sense of the distribution of sources of ED referrals, we use the same merged dataset and obtain a count of each of the entry types under the "Admission_Source" field (originating from EDIS prior to the merge). The resulting distribution is shown in Figure 3-5, where we notice that the majority of ED referrals (~84.7%) originate as patient self-referrals/walk-ins (note that we consider 'EMS transport decision' as a patient self-referral too).

[17] These 3,801 hospital admission encounters correspond to 2,031 unique patients. In other words, there are cases where more than one hospital admission is from the same patient.

[18] Note that "timestamps" were extracted from the date and time strings to account for the (relatively) short time elapsed during patient transfer from ED to hospital.

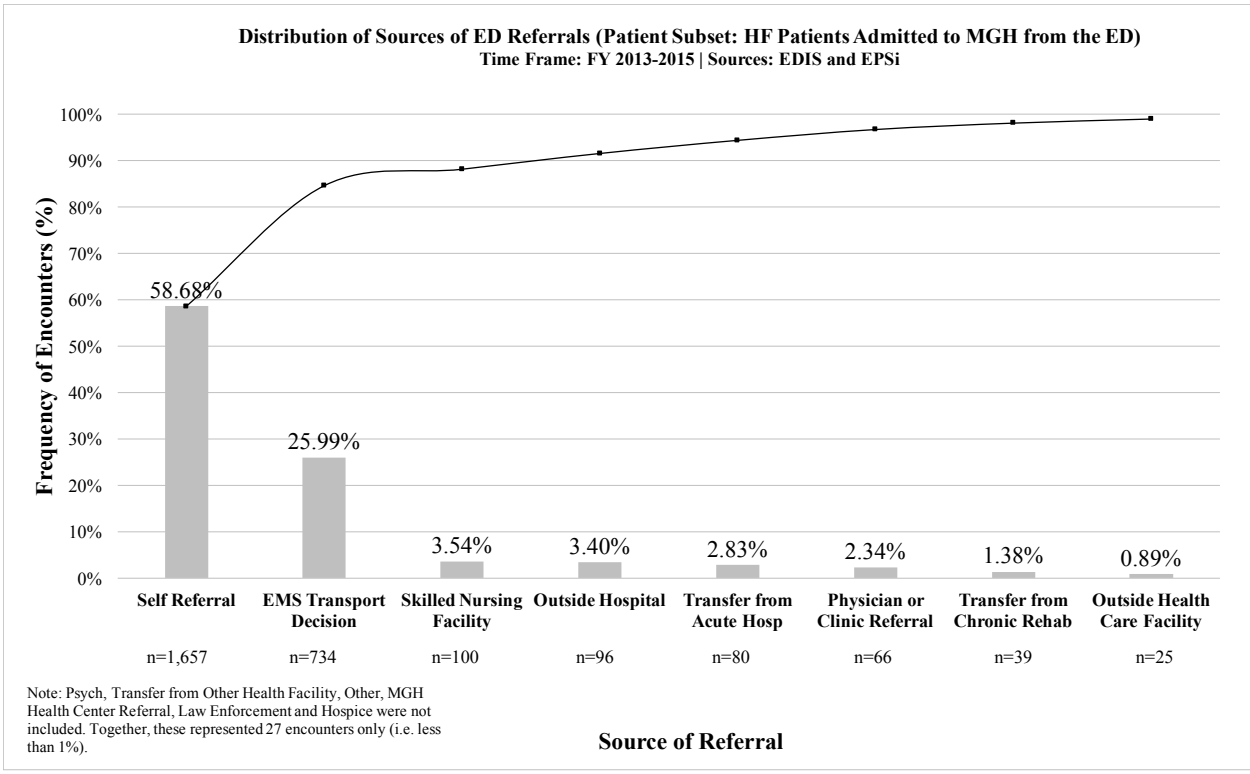


Figure 3-5: Distribution of Sources of Emergency Department Referrals; note that 'EMS Transport Decision' is also considered 'Self-Referral'

3.5 Inpatient Care during Heart Failure Hospitalizations

As mentioned previously, an admitted patient can be admitted to the hospital's medicine service or the cardiology service. Depending on the service and care provider, the patient would undergo a heart failure assessment, diagnostic testing, treatments (e.g., fluid management), and education on how to best manage heart failure (e.g., set targets for weights and fluids, discuss a healthy diet plan, etc.). If the patient has a known cardiologist on record, he/she may be reached out to as part of the assessment. However, it should be noted that communications with other providers for inpatient consultations is inconsistent and some attending physicians in the hospital may choose not to follow-up with members of the outpatient cardiology care team.

Just prior to the hospital discharge, an outpatient follow-up visit should be scheduled with a member of the cardiology care team. At this point, if the admitted patient had family members present with him/her, the medical team in the hospital would also discuss patient's care plan with them. Finally, the patient's discharge note is documented with provider contact information and instructions regarding prescribed therapy at discharge. The inpatient care pathway is shown in Figure 3-6.

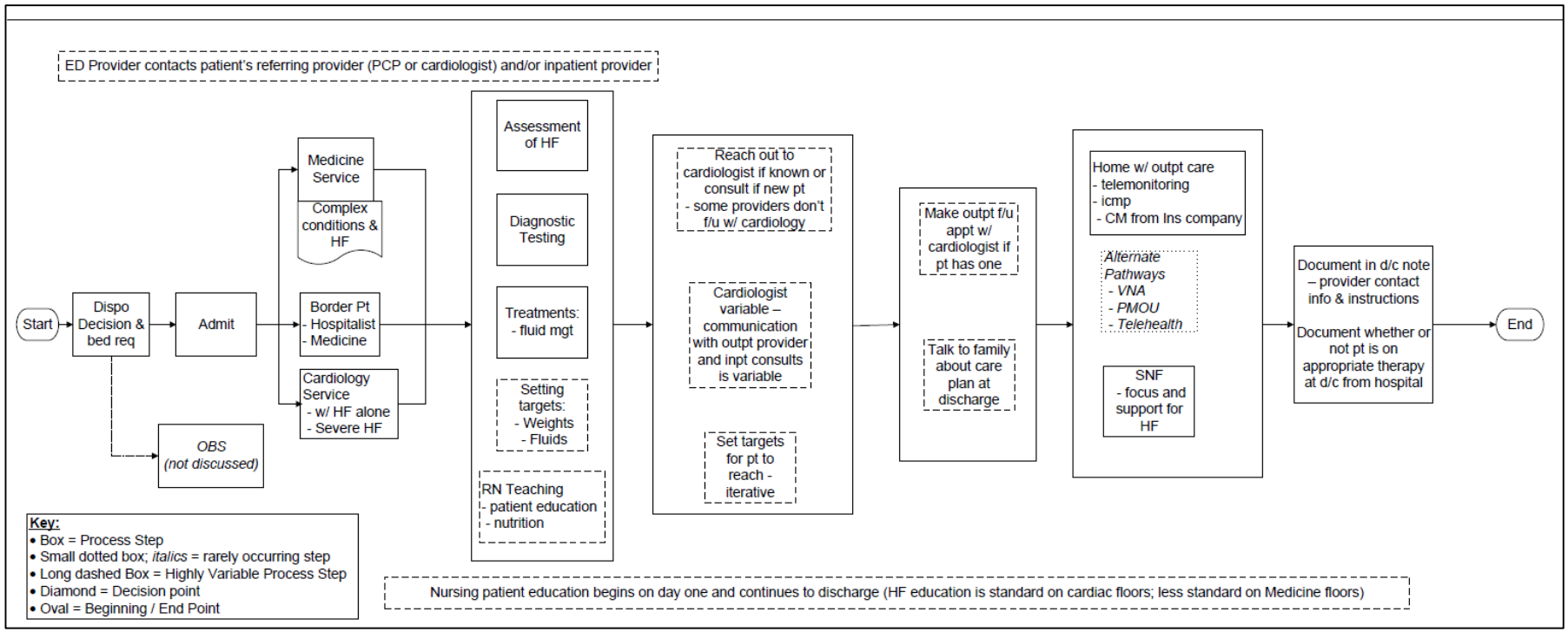


Figure 3-6: Inpatient Heart Failure Care Pathway

3.6 Post-Discharge Outpatient Care, Monitoring and Escalation

Similar to the ED, there are several discharge dispositions that are used with heart failure patients post hospitalization. These are primarily determined by the severity of the disease. The five main dispositions at MGH are as follows: (i) discharge with no home services; (ii) discharge with home services (e.g., VNA care); (iii) discharge to a skilled nursing facility (SNF)[19] or rehabilitation center; (iv) discharge with integrated care management program (iCMP) and telemonitoring; and (v) discharge under hospice. The discharge disposition that a patient is assigned to depends on multiple factors including patient preferences and the individual physician's decision on whether or not home services are needed.

The various possible discharge dispositions are a source of variability in care among different heart failure patients. First, a discharge with some form of services or extended care beyond the inpatient setting is often intermittent in nature. For example, in the case of iCMP (refer to Section 2.3.2 for more details), patients might be enrolled in the program for a while before leaving it and re-entering at some other point. As such, the transient nature of enrollment and the variation in length of enrollment can have an impact on the time that a patient's next admission will take place. Second, care and medication management is not standardized across all types of discharge services. For example, for patients receiving care from a member of the visiting nurse association (VNA), the degree of communication between the VNA and PCP/specialist is variable and different than patients discharged into a SNF. In fact, even among different SNFs, patients might experience varying levels of monitoring, communication with their PCP/cardiologist, and escalation where needed.

[19] Refer to Appendix A for definitions.

Telemonitoring is also available as a post-discharge service. However, as discussed earlier in Chapter 2, it also has several limitations that affect its success in monitoring patients. First, the telemonitoring that MGH employs involves monitoring patients remotely through a number of monitoring devices at home, with the patient indicators sent via telephone to the health care provider. Often, the physiological parameters being monitored (e.g., mainly weight in the case of MGH) are not very effective at accurately detecting decompensating heart failure[20]. Furthermore, telemonitoring has proven useful only when patients and providers are compliant with recording and monitoring results, respectively. Also, it is most successful at the earlier stages (i.e., I and II) of heart failure disease.

3.7 Generating Hypotheses around Heart Failure Care

As an outcome of the mapping activities and discussions, we present two key hypotheses around the management of heart failure patients. These include: (i) *heart failure patients face serious access problems to outpatient care services, including PCPs and cardiologists*, and (ii) *there is a potential to predict which patients are at the greatest risk to be hospitalized for the first time because of a heart failure condition.*

[20] Patients at risk for heart failure hospitalization often suffer from an accumulation of fluid volume that leads to an increase in cardiac filling pressures. Because changes in volume are often apparent several weeks before symptoms worsen, the availability of an accurate and responsive measure of a patient's fluid volume status is essential. Unfortunately, body weight and volume status 'diverge' with time from hospital discharge, mainly due to target dry weight changing in response to calorie intake. Although, on average, patients do gain weight before hospitalization for decompensation, amount of weight gain is often fairly modest. Thus, rapid weight gain is considered a "relatively poor surrogate" for volume status and, consequently, cardiac filling pressures [44].

In Chapter 4, we take a closer look at the first hypothesis, what motivated it, and the analyses performed to validate it. Upon confirming the seemingly lacking access to outpatient care, we then turn to the second hypothesis and discuss the approach taken to confirm the predictability claim. This is discussed in Chapter 5.

4. Current State Analysis: A Quantitative Approach

4.1 Overview of Analysis Methods and Approaches

Up to this point, we followed a more qualitative approach to describe the MGH care system for heart failure patients and the problems that might exist. We now turn to quantitative analyses that serve the purpose of validating (or rejecting) the hypotheses presented at the end of Chapter 3. In this chapter, we focus primarily on the outpatient access problem. Specifically, in Section 4.2.1, we focus on patients admitted to the hospital with a heart failure condition and take a closer look at whether they had scheduled outpatient appointments with primary care providers and/or cardiologists over the weeks leading up to their admissions, as well as following their hospital discharges. In addition, in Section 4.2.2, we analyze the appointment scheduling lead times for those patients who did have completed appointments.

4.2 Identification and Quantification of Gaps in the Current Care Structure

4.2.1 Outpatient Interaction Prior to Hospitalizations and Post Discharge

The finding that most ED visits originate as self-referrals (described earlier in Section 3.4.2) led us to question the access that heart failure patients have to outpatient care and, hence, develop the first hypothesis we presented in Section 3.7.

As a first pass towards validating our hypothesis, we turn to an analysis of the distribution of ED arrival times, for patients admitted to the hospital following an ED visit. First, we consider the frequency of ED visits for each day of the week, using the same dataset and time frame as the one described in Section 3.4.2 to develop a distribution of patient sources of

admission. Second, we analyze the frequency of ED visits based on time of occurrence, using 3-hour time intervals. The results, shown in Figure 4-1 and Figure 4-2, respectively, indicate that approximately 77.3% of hospital admissions via the ED occur during weekdays and about 68.0% between the hours of 9am and 6pm.

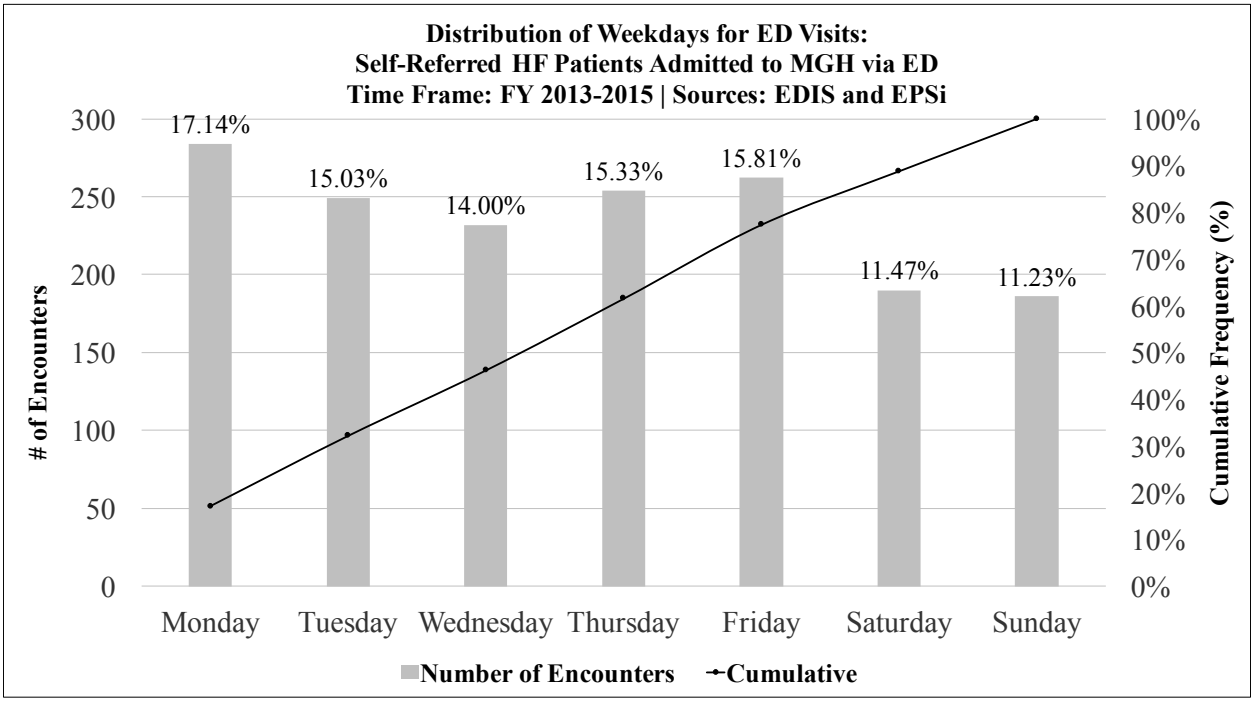


Figure 4-1: Distribution of Weekdays for Self-Referred Heart Failure Patients Admitted to MGH via the ED

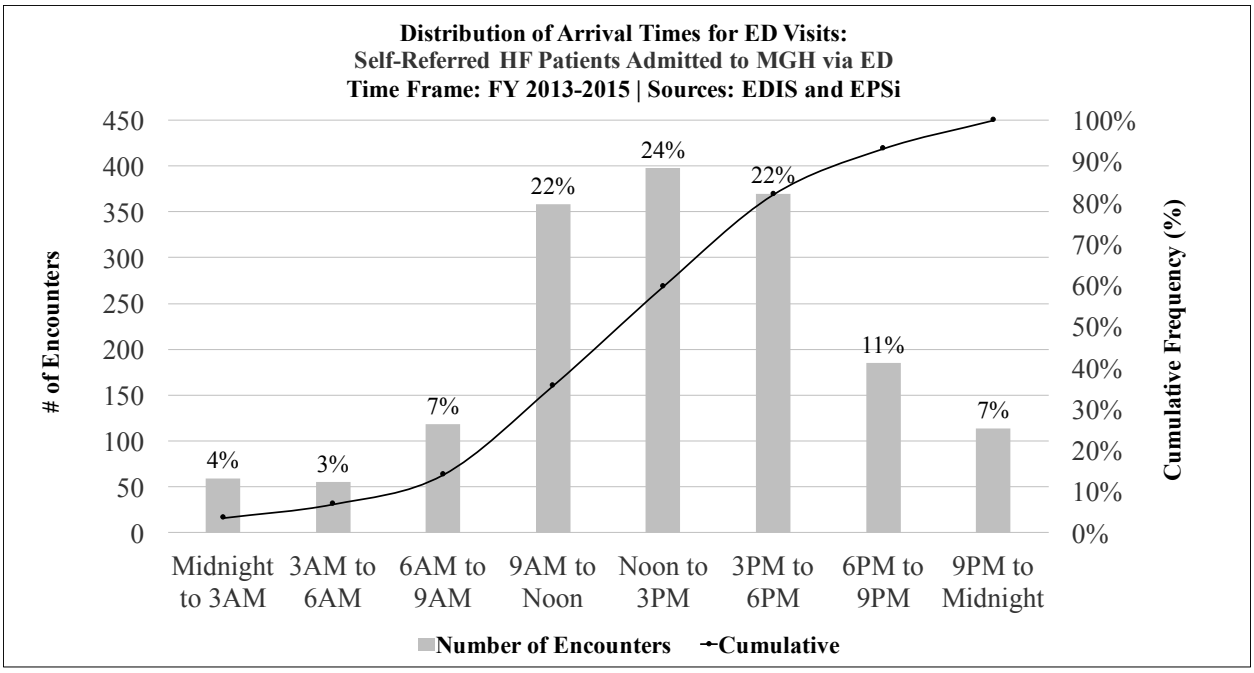


Figure 4-2: Distribution of ED Arrival Times for Self-Referred Heart Failure Patients Admitted to MGH via the ED

The finding that the majority of heart failure ED visits leading to hospital admissions occurred during weekdays and between the hours of 9 am and 6 pm prompted further analysis of the access to outpatient appointments. In particular, clinical input suggests that the two weeks prior to a heart failure hospitalization often represent a time period where the detection of heart failure decompensation signs is likely, thereby permitting early and corrective therapeutic intervention that could ultimately prevent the need for hospital care. Similarly, outpatient follow-up appointments in the eight weeks post discharge is important to monitor clinical stability and ensure that patients are taking the proper and necessary steps for disease management. The hypothesis that motivated the following analyses was that heart failure patients likely face an outpatient access problem that seems to be causing them to effectively treat the ED similar to a primary care or outpatient clinic (i.e., in terms of their times of arrival/visits).

To quantify outpatient access, the number of scheduled and completed outpatient appointments with primary care physicians and cardiologists was used as a metric. Hospital scheduling data, stored in the Cadence Scheduling Database, was used to access scheduled appointments in a fairly straightforward process. For each heart failure patient, the date of a hospital admission with a heart failure primary discharge code was obtained. Using this date, outpatient appointments with cardiology[21] and primary care clinics[22] in the two weeks leading up to a hospitalization (or eight weeks post discharge) were then identified. These appointments were classified as completed, canceled, or no-show. Patients that did not have an appointment in any of these categories were classified as not scheduled. Finally, heart failure patients were divided into groups based on their institutional ownership. The groupings were the same as those used in the classification of heart failure patients (discussed in Section 3.3.1).

The results, presented in Figure 4-3 and Figure 4-4, show that a very large fraction of patients are not even scheduled. Specifically, out of a total of 1,115 admitted heart failure patients in FY2015, about 56.7% of patients did not have a scheduled outpatient appointment in the two weeks prior to their hospitalizations. Similarly, about 42.9% of the admitted patients did not have scheduled follow-up appointments in the eight weeks following their discharge.

We draw attention to two points in Figure 4-3 and Figure 4-4. First, when canceled and no-show appointments are also considered, we observe even higher fractions of patients with no completed appointments (prior to hospitalizations and post discharge). Second, we draw extra attention to the outpatient appointments of patients owned by MGH PCPs (i.e., Groups 1 and 2) since these are a priori more likely to have appointments, in addition to the fact that MGH could

[21] Cardiology clinics included Cardiology, Cardiac Rehabilitation and Cardiac Surgery

[22] Primary Care clinics included Primary Internal Medicine and Family Medicine

intervene more on them. On average, 56.4% and 55.6% of PCP-owned patients have no completed appointments in the two weeks prior to hospitalization and eight weeks post discharge, respectively.

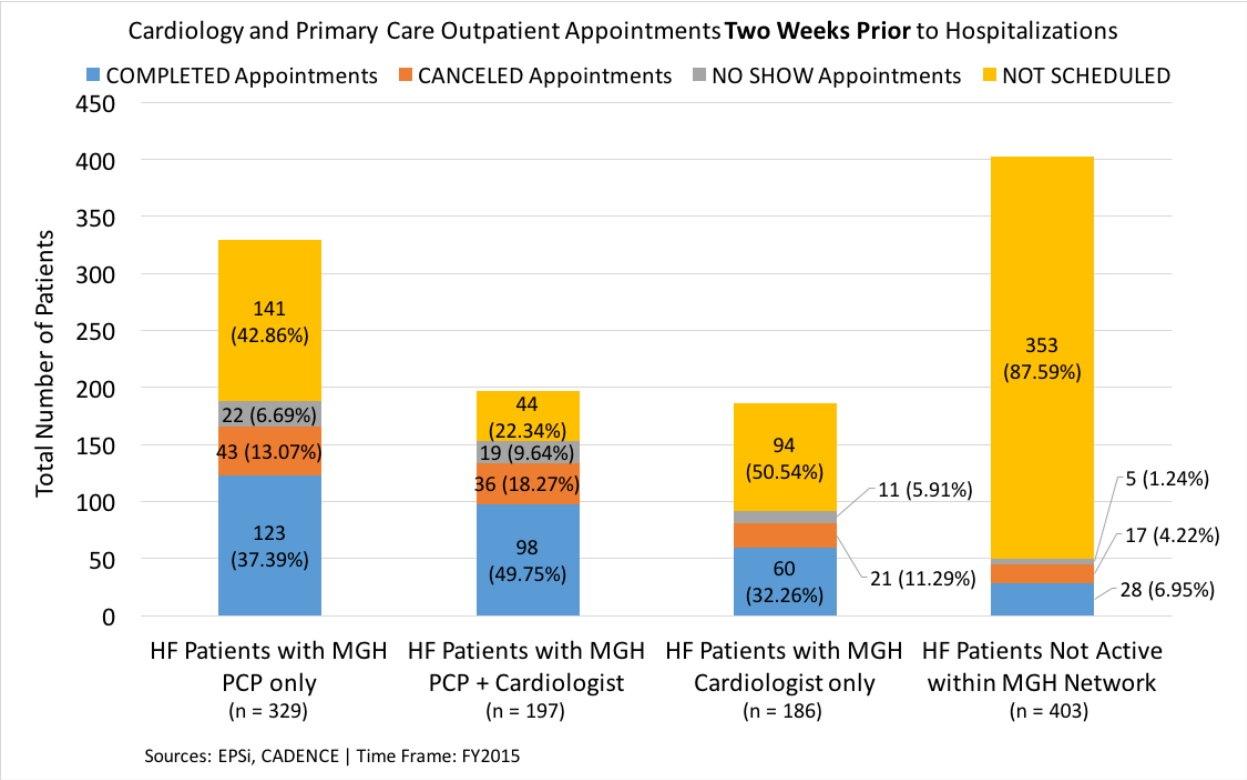


Figure 4-3 Cardiology & Primary Care Outpatient Appointments Two Weeks Prior to Hospitalizations

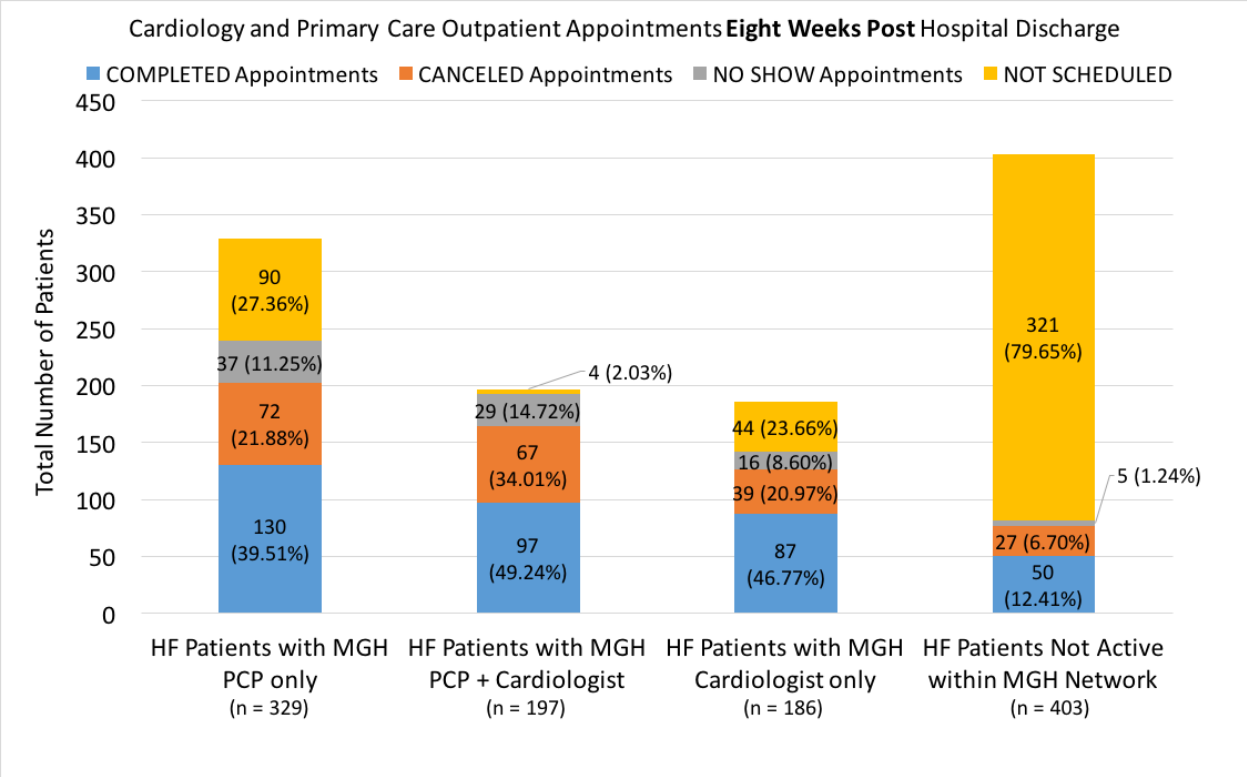


Figure 4-4: Cardiology & Primary Care Outpatient Appointments Eight Weeks Post Hospital Discharge

4.2.2 Outpatient Appointment Lead Times

To better understand the outpatient access, completed outpatient appointments are further analyzed, with a focus on appointments just prior to hospitalization events. As shown in Figure 4-5, completed outpatient appointments are first divided into two groups, depending on whether the appointment was with a cardiology clinic or primary care office. Within each category, completed appointments are then grouped based on the number of days elapsed from the outpatient appointment date to the first admission event that took place (following the appointment). For each of these subcategories, we obtain the distribution of appointment lead times, calculated as the difference in time between the appointment booking date and the actual appointment date.

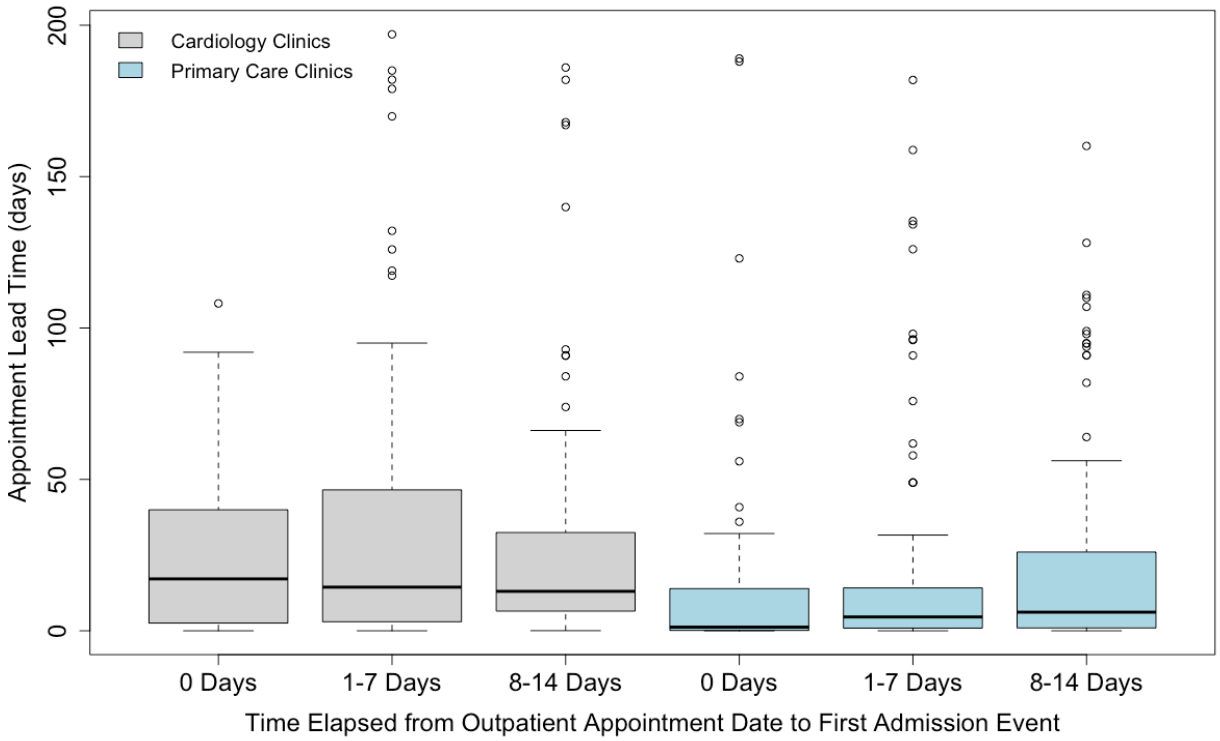


Figure 4-5: Distributions of Scheduled Appointment Lead Times for Cardiology and Primary Care Clinics; Source: EPIC and Cadence; Time Frame: Oct '14 – Sep '15

There are two main insights from the analysis of appointment lead times. First, as Figure 4-5 shows, there are patients whose appointment lead times are very long, especially when cardiology clinics are considered. More specifically, median appointment lead times lie in the range of about 13 to 17 days, which exceeds the critical two-week time period of heart failure decompensation prior to a hospitalization. To make matters worse, patients grouped under the 'bucket' of zero elapsed days from the cardiology clinic outpatient appointment date to first admission event (i.e., the first cardiology boxplot in Figure 4-5) are the ones with the highest median appointment lead time. This suggests the hypothesis that patients are attempting to schedule appointments but, because of the long lead times, end up being sent to the ED upon arrival for their appointments. The result is a failure to detect and control gradual deteriorations

in clinical stability before the patient arrives at the ED. As such, it represents a missed opportunity to intervene in a timely manner to keep the patient out of the hospital.

Second, cardiology clinics face more serious outpatient access problems than primary care clinics, as shown by their higher median appointment lead times. This presents an issue for a condition like heart failure, which, due to its progressive nature, may require specialized care teams capable of effectively carrying out the necessary disease management for more complex stages. The lack of a heart failure patient registry to monitor and track these patients, as well as ensure they receive timely follow-ups, seems to contributing to this serious problem.

5. Predictive Modeling of Heart Failure Patients' Hospitalization Risk at MGH

5.1 Objective and Overview

The analyses and findings presented thus far helped in generating an understanding of the current state of the heart failure care pathway, the volumes of heart failure ED visits and hospital admissions, and the need for increased outpatient access. However, implementing more targeted interventions requires the design of predictive models that aim at identifying patients that are at the greatest risk of being hospitalized.

Our decision to focus on predicting first-time admissions is motivated by findings from previous work that identified previous inpatient bed days as one of the most significant predictors of future admissions [51]. As such, a way to prevent readmissions is to prevent the first admission to begin with. This goes back to the earlier discussion in Section 1.2.3 where we find that the first decompensation and, consequently, hospitalization event results in the greatest drop in clinical stability. Thus, preventing this significant dip in stability, and the patient's entry to an active and risky state of disease progression, is important.

Figure 5-1 illustrates the problem to be solved. Assuming that it is the end of a given year[23] *t*, the first step is to identify the patients that only had records of outpatient (OP) encounters in year *t* and all the prior years[24]. In other words, any patient with an inpatient (IP) encounter prior to and/or during year *t* is excluded. We then further limit this group of patients,

[23] In our modeling, we focus on two specific scenarios for predicting first-time heart failure admissions. They take place at the end of years (i) $t = 2013$ and (ii) $t = 2014$.

[24] Note that were are limited to data from January 2011, which serves as our starting point.

with an outpatient appearance only, to those patients affiliated with an MGH primary care provider. This step is carried out to eliminate any noise that might affect the models as a result of, say, one-off visits by patients that do not receive longitudinal care within the MGH network[25]. Furthermore, we do not have complete data for patients that are not directly within the MGH network.

The population of interest is marked (shaded in gray) in Figure 5-1. Out of this population, we attempt to predict the patients that will have their first admission in the following year, t + 1. As we shall see later in Table 5-2, these first-time admissions represent a minority event for outpatients. On average, 6.78% of heart failure patients with outpatient encounters only are expected to have a first-time hospital admission.

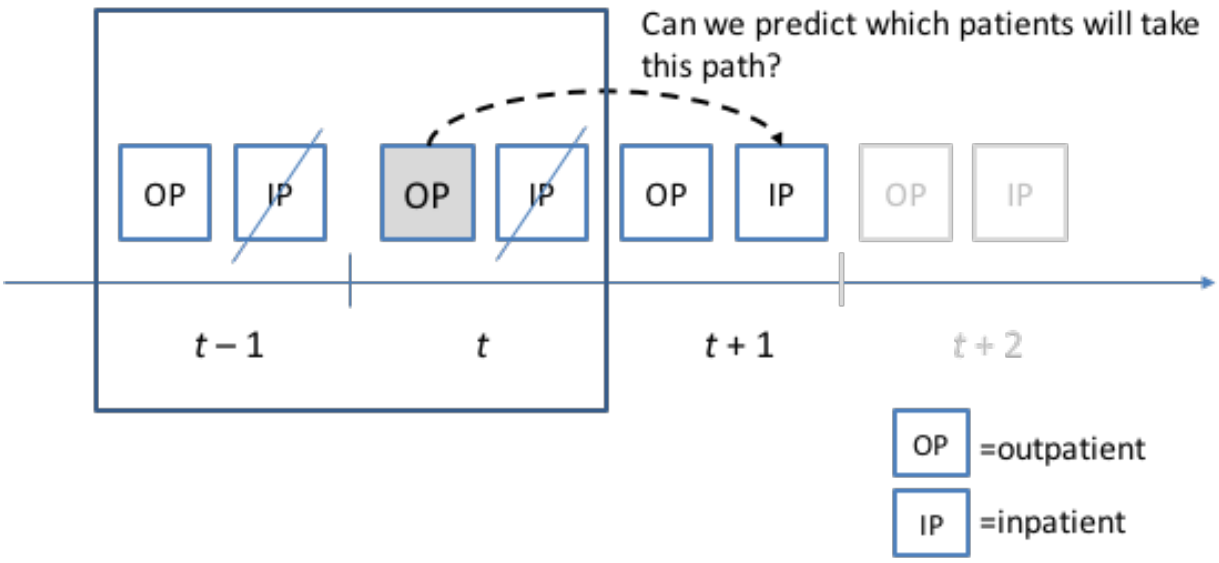


Figure 5-1: Schematic of Heart Failure Population of Interest for First-Time Hospital Admission Predictions

[25] As mentioned earlier, in this exercise, we only analyze inpatient and outpatient encounters within the MGH network. This analysis could also be extended to include other Partners facilities. However, including other encounters outside the Partners network would be more difficult due to data availability.

5.2 *Risk Factors and Predictive Features*

Generating a list of risk factors is the first step in any predictive modeling exercise. The aim was to develop a list of features that are hypothesized to be predictive of heart failure patients' admissions or, more specifically, their first-time admissions.

To generate such a list, the existing literature was reviewed to understand similar predictive models and the corresponding risk factors they used[26]. This was helpful to identify the important general patient characteristics and demographics that should be included. Examples of such factors include gender, age, marital status and ability to speak English. Moreover, the clinical literature was helpful in identifying clinical indicators that are of relevance and these include: systolic blood pressure (BP), pulse and ejection fraction (EF)[27]. Gender, marital status and spoken language were modeled as binary factors. Age, systolic blood pressure, pulse and ejection fraction were modeled as continuous numerical variables. For blood pressure, pulse and ejection fraction, two types of variables were included. First, the patient's worst reported measurement of each was included. Second, in an attempt to account for trends and changes over time in each of these variables, the percent change between any two consecutive measurements was calculated for each variable. The overall average percent change for each patient was then used as the second clinical predictor.

The remaining independent variables that were used as features in the model were identified based on regular meetings and interactions with clinicians including cardiologists,

[26] The primary difference between our model and those reported in the existing literature was the outcome variable. The models we consulted for insight on predictive features were focused on predicting 30-day readmissions while our model aims to predict first-time heart failure admissions.

[27] Refer to Appendix C for a glossary of the clinical predictors used in our model.

primary care providers and nurses. As a first group of variables, these included administration of various cardiac medications, some of which were hypothesized to be associated with better outcomes (specifically, fewer hospital admissions) and others with worse outcomes (or more hospital admissions). The former included the following classes of drugs: beta-blockers, angiotensin-converting-enzyme (ACE) inhibitors, angiotensin receptor blockers (ARBs) and aldosterone antagonists. The latter included loop diuretics. Within each class, the generic names of medications were also provided. The drugs that were part of this project's analysis are shown in Table 5-1, together with the class they belonged to. In addition to accounting for the presence, or lack of, heart failure specific medications, the total number of medications administered and/or prescribed to a patient was also included as a proxy for number of comorbidities and, thus, a predictor of a patient's clinical complexity.

Table 5-1: Evidence-Based Heart Failure Drugs and Hypothesized Effects on Hospital Admissions

Drug	**Hypothesized Effect on Hospital Outcomes (specifically, hospital admissions)**
Beta-blocker	
Metoprolol	Associated with fewer hospital admissions
Carvedilol	Associated with fewer hospital admissions
ARB	
Valsartan	Associated with fewer hospital admissions
Aldosterone antagonist	
Spironolactone	Associated with fewer hospital admissions
Loop Diuretic	
Furosemide	Associated with more hospital admissions
Torsemide	Associated with more hospital admissions
Bumetanide	Associated with more hospital admissions

Second, as a measure of cognitive abilities, the presence of a history of psychotic medication prescriptions was included. Based on clinical input, a list of generic psychotic medication names was developed, which is presented in Appendix D. In the case of both heart failure drugs and psychotic medications, variables were modeled as binary factors indicating the presence or lack of respective medications in the patient's medical history within the study's time frame.

In addition to variables representing clinical symptoms and medications, including variables representing patient ownership and engagement was deemed important too. Here, variables like the total number of any outpatient visits (from 2011 to time of prediction), the total number of evaluation and management (E&M) visits with cardiologists specifically, and the time elapsed since a patient's first diagnosis were included and modeled as continuous numerical variables. Moreover, since this analysis was limited to patients that are affiliated with MGH primary care providers, institutional ownership through PCP linkage, though not directly modeled as a predictor variable, was also a controlled feature in this study.

Finally, socioeconomic factors were also included as part of the study due to heart failure's large dependency on patient lifestyles. For example, this included a patient's income, estimated using median incomes reported by the U.S. Census Bureau for all U.S. zip codes[28] and modeled as a continuous variable [52]. In addition, any history of substance abuse was also identified using ICD-9 and ICD-10 codes. Similar to psychotic medications, this variable was also modeled as a binary input. Details on the specific codes used are presented in Appendix E. A detailed summary of all predictor variables, along with their respective names in model summaries and expected coefficient signs, is given in Appendix F.

[28] Patient zip codes were obtained from EPSi

5.3 Descriptive Statistics and Data Preparation

5.3.1 Developing Year-by-Year Datasets

With a list of predictive features, we now proceed to develop three separate datasets of patients, each corresponding to patients that only had outpatient encounters with heart failure diagnosis codes in 2011-2012[29], 2013 and 2014. Then, for each patient that appeared with a heart failure code in 2011-2012, predictor variable data is added from the same time frame. Additionally, for each patient, a "1" or "0" is assigned as the output dependent variable corresponding to the presence of a first-time admission in the following year, 2013, or the lack of one, respectively. The same approach is applied for outpatients appearing in 2013 and 2014, but with first-time admissions being identified in 2014 and 2015, respectively. Figure 5-2 shows schematics of these datasets. Note that first-time admissions in year 't + 1' represent the binary dependent variable of a dataset from year 't'.

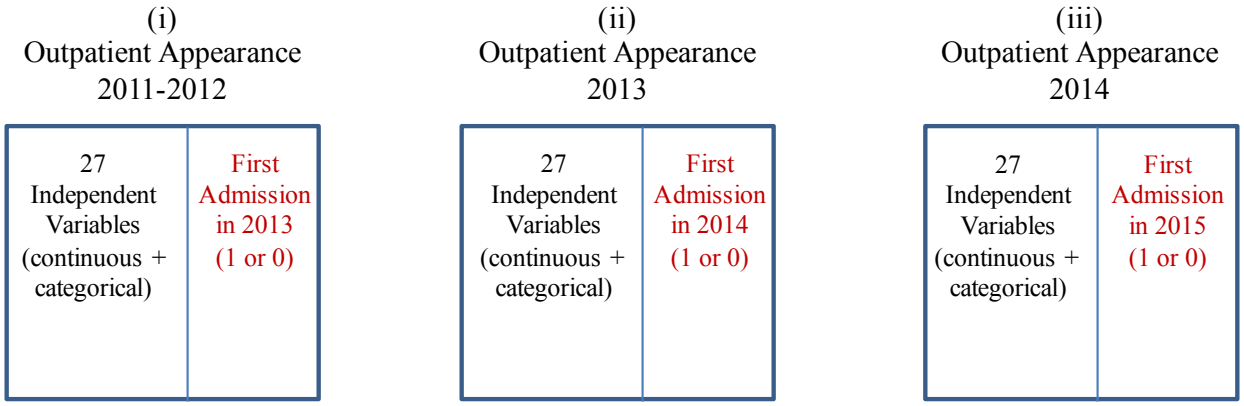


Figure 5-2: Representation of Year-by-Year Model Datasets

The numbers of patients from each dataset are shown in Table 5-2. It should be noted that, in this table, each row represents patients that are necessarily a subset of the preceding row

[29] Due to the unavailability of data for the entire fiscal year of 2011, data from 2011 was grouped with that of 2012

(except for row #1). Clearly, the number of patients with a first-time admission represent a minority of the datasets, which points to a relatively low 'rate of conversion' of heart failure patients from the outpatient to the inpatient setting[30].

Table 5-2: Numbers of Heart Failure Patients in 2011-2012, 2013 and 2014 Datasets

		Jan '11 – Sep '12 (2011–2012)	Oct '12 – Sep '13 (FY 2013)	Oct '13 – Sep '14 (FY 2014)
1	# Patients with any Outpatient Heart Failure Diagnosis Appearance	3,352	4,954	5,677
2	# Patients with no Inpatient Admission	2,583	3,406	4,157
3	# Patients Affiliated with MGH PCP	1,249	1,380	1,577
4	# Patients with a First-Time Admission in the Following Year	104	93	83
5	Fraction of MGH PCP Affiliated Patients with a First-Time Admission (i.e., '**4**'/'**3**'*100%)	8.33%	6.74%	5.26%

5.3.2 Developing an Overall Dataset and Oversampling Minority Events

To develop our model, the year-by-year datasets described in Section 5.3.1 are combined (using R's 'rbind' function) to form one single dataset of 4,118 rows, as shown in Figure 5-3. In this larger and combined dataset, the fraction of first-time admissions is 6.65% (which is close to the overall average of 6.78% from the three individual datasets reported in Table 5-2).

Next, we use the 'mice' function of the 'mice' package in R to address any missing data in the larger training set. The 'mice' function works by generating multiple imputations for

[30] Note that the percentages of first-time admissions (i.e., 8.33% in 2011-2012; 6.74% in FY2013; and 5.26% in FY2014) are not statistically different across the years.

incomplete multivariate data by Gibbs sampling[31]. The algorithm imputes each incomplete (or 'target') column by generating 'plausible' synthetic values given all other columns in the data[32] [53].

To avoid misclassifications that might arise due to training an over-conservative model that predicts "no first-time admission" most (or all) of the time, we turn to techniques like oversampling minority events (in this case, presence of first-time admissions). The aim here is to improve the predictions through more accurate classifications. To achieve this, first-time admissions are increased to represent 50%[33] of the training set. Out-of-sample test sets are left unchanged (i.e., no oversampling is applied to them). The 'ubOver' function of the 'unbalanced' package in R is used to perform the oversampling. Through sampling with replacement, this function generates random replicates of some instances from the minority class of a dataset in order to obtain a final dataset with specified fractions of instances for the different classes.

[31] Gibbs sampling is a statistical technique used for obtaining a sequence of observations which are approximated from a multivariate probability distribution. It is a randomized algorithm that generates a chain of samples, each of which is correlated with other nearby samples.

[32] For predictors that are incomplete themselves, the most recently generated imputations are used to complete the predictors prior to imputation of the target column.

[33] In the literature, oversampling minority events is recommended up to levels of 10% - 50%. For this project, this entire range was explored, but the out-of-sample AUC was found to change very little over the range. Since initial analyses were performed with 50% oversampling, and little to no change was observed with lower percentages, we kept a threshold of 50% oversampling for all regressions.

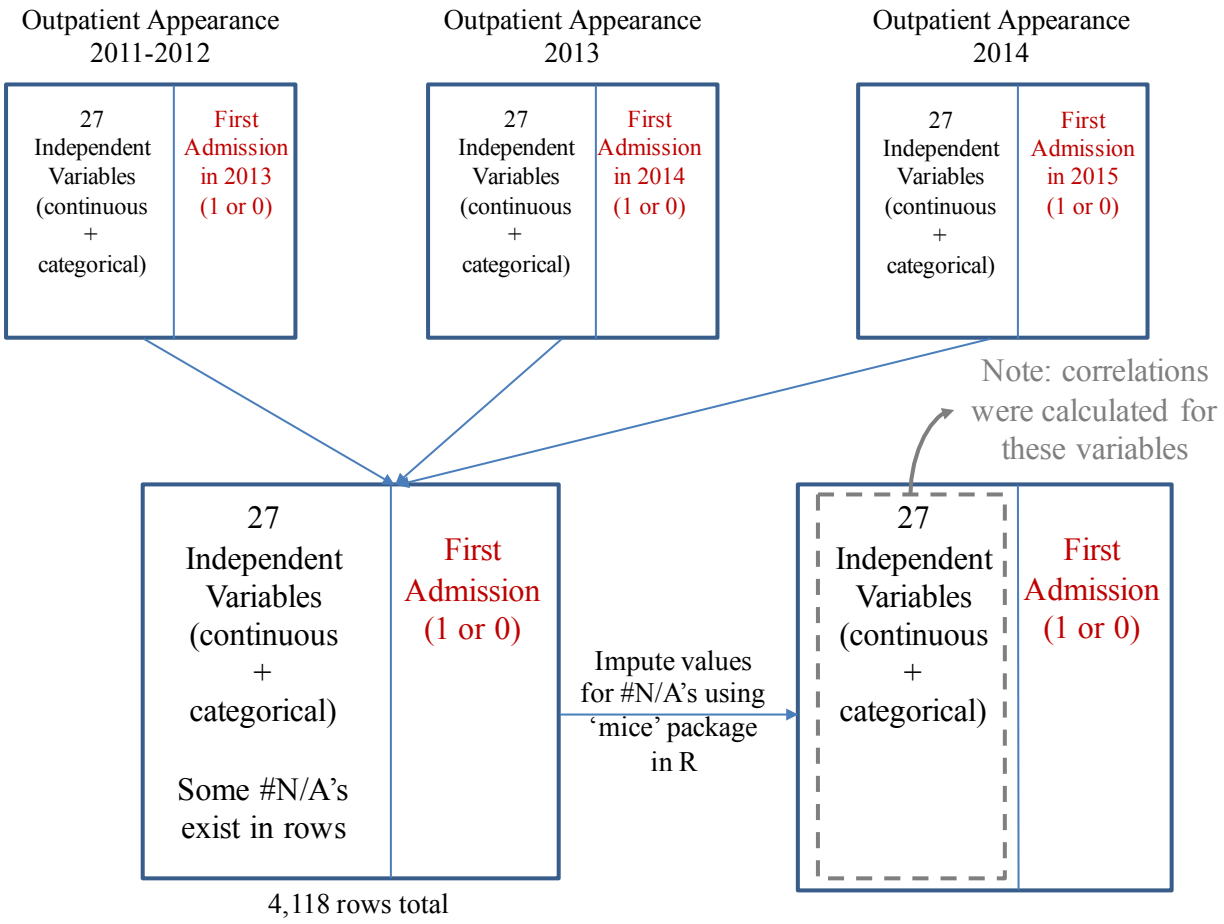


Figure 5-3: Methodology of Deriving the Final Dataset for Model Training and Testing

5.3.3 Calculating Correlations and Eliminating Multicollinearity between Predictor Variables

To calculate correlations between predictor variables, we run R's 'cor' function on a design (or model) matrix of the final dataset's independent variables. The results are shown as a heat map in Figure 5-4. As a threshold, we take correlations that are greater than 0.3 or less than -0.3 as 'high' and drop a variable out of highly-correlated pairs.

Overall, 10 variables are dropped. These include: mean systolic blood pressure, average percent change in systolic blood pressure, maximum pulse, mean pulse, minimum ejection

fraction, mean ejection fraction, administration of carvedilol, administration of valsartan, administration of torsemide, and presence of a history of substance abuse. The reasoning behind dropping the first six variables is straightforward and intuitive. Here, we notice that clinical indicators are measuring the same physical characteristic and then reporting different statistics of the same variable (i.e., mean, minimum and average percent change). It is therefore expected that these will be correlated. In the remaining four variables, the reasoning behind dropping variables is not as straightforward. In this case, we hypothesize that some patients might be taking more than one type of drug at any time. In turn, this might mean that there are some pairwise interaction terms that still need to be accounted for. However, for the sake of simplicity and model interpretability at this stage, we drop the 'troublesome' variables and reduce our medication variables to two medications associated with 'good outcomes' and another two associated with 'worse outcomes' (fewer and more hospital admissions, respectively).

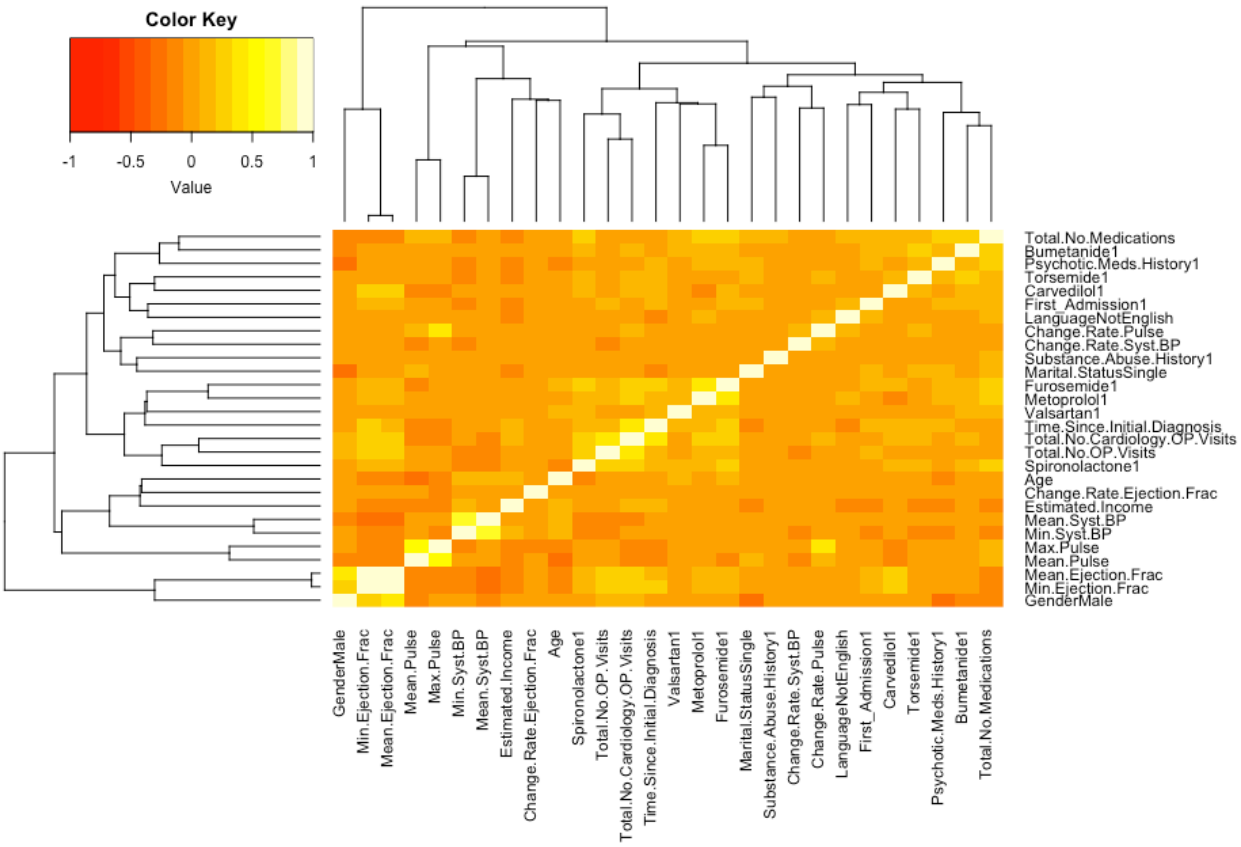


Figure 5-4: Heat Map of Correlations between Predictor Variables

5.3.4 Training/Test Partitions and Modeling Techniques

With the overall dataset now reduced to only 17 predictor variables, we next use the 'createDataPartition' function of R's 'caret' package to create a series of stratified, random 80/20 training/test partitions. More specifically, the function randomly splits the 4,118 rows into a training set that represents 80% of the source dataset and a test set that represents the remaining 20%, all while keeping the fraction of first-time admissions close to 6.65% in both cases. Multiple (specifically 15) iterations of splits are carried out and, each time, the training set is oversampled so that first-time admissions are increased from about 6.65% to 50%, while the test set is kept unchanged.

For each oversampled training partition, three different types of models are then developed and each is tested for its out-of-sample performance using the corresponding test partition. The three modeling techniques we use include: (i) logistic regression; (ii) a regression trees; and (iii) random forests. To settle on a final modeling technique, we compare the different models on their overall, average out-of-sample performance, using average area under the receiver operating curve (AUC) as the evaluation metric (described in Section 0).

As we shall see in Section 5.4.4, logistic regression outperforms regression trees and random forests in out-of-sample tests. As such, we base the following sections solely on the logistic regression technique and return to the two other techniques later to compare model performance.

5.4 Developing and Evaluating a Final Predictive Model

5.4.1 Logistic Regression: Overview and Evaluation Method

Logistic regression is a useful modeling technique in applications where the dependent variable is categorical. This is a problem that is often encountered in healthcare settings, where some prediction or evaluation needs to be made on a large scale and has a discrete nature (either binary or multiple, ordered categories). For example, the Trauma and Injury Severity Score (TRISS), which is widely used to predict mortality in injured patients, was originally developed by Boyd et al. using logistic regression [54]. Similarly, logistic regression is seen in applications that predict whether a patient has a given disease (e.g., diabetes or coronary heart disease) based on observed patient characteristics like age, sex, blood test results, etc [55].

In this project, we use logistic regression as an analytic tool to predict whether a patient will be admitted for the first time in a given year or not, and thereby guide future care

interventions. More specifically, logistic regression does this by predicting the probability that a first-time admission will take place using the logistic response function:

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}}$$

where,

y: dependent variable 'first-time admission'

$x_1, x_2, \ldots, x_k$: identified predictors or independent variables

$\beta_0, \beta_1, \ldots, \beta_k$: regression coefficients

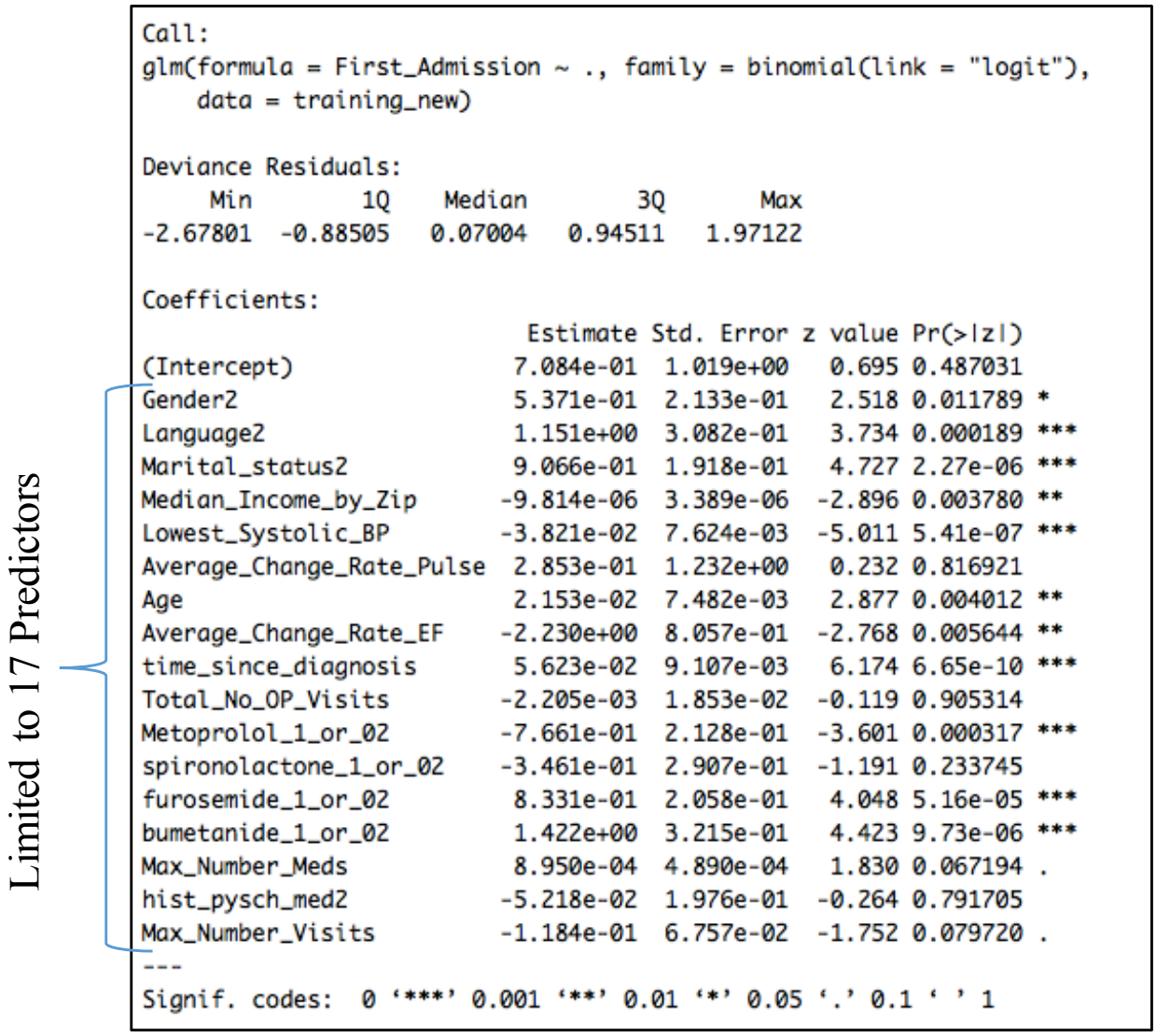
The logistic regression coefficients are estimated using a maximum-likelihood estimation. The aim is to select the parameters of the statistical model in such a way as to predict a high probability for the first-time admission cases and a low probability for the cases of no first-time admissions.

In order to translate the probability outcome of a logistic regression model into a class prediction that can be compared with the actual outcome of presence or lack of a first-time admission, a threshold value m is selected. Then, if $P(y=1) \geq m$, presence of a first-time admission (i.e., output "1") is predicted. Conversely, if $P(y=1) < m$, lack of a first-time admission (or output "0") is predicted. Threshold values are then selected based on the errors associated with "lower costs". In this case, since admissions are being predicted to guide proactive outpatient care and reach-out on the hospital's end, erring on the conservative end is deemed 'better'. In other words, predicting that an admission will happen, when in reality it did not actually happen, is "better" than predicting that an admission will not happen and then discovering that it did take place. To enable such a scenario, m is preferred to be small (or less than 0.5).

5.4.2 Identifying Significant Variables through the Akaike Information Criterion

As described in Section 5.3.4, we split our dataset into 15 partitions and run a logistic regression on each of them, as one form of modeling technique. For each partition and its corresponding logistic regression run, we perform model selection using the Akaike information criterion (AIC)[34] and record the variables that appear as most significant. Figure 5-5 shows one example of the output summary of a regression run. All 15 runs are presented together in Appendix G.

[34] AIC is a measure of the relative quality of statistical models for a given set of data and parameters. Given a collection of models for the data (i.e., with different parameters included in each), AIC provides a relative estimate of the 'information lost' when a certain model is used to represent the process that generates the data. In doing so, it offers a trade-off between the goodness of fit of the model and its complexity.




```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.67801  -0.88505   0.07004   0.94511   1.97122

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                7.084e-01  1.019e+00   0.695 0.487031
Gender2                    5.371e-01  2.133e-01   2.518 0.011789 *
Language2                  1.151e+00  3.082e-01   3.734 0.000189 ***
Marital_status2            9.066e-01  1.918e-01   4.727 2.27e-06 ***
Median_Income_by_Zip      -9.814e-06  3.389e-06  -2.896 0.003780 **
Lowest_Systolic_BP        -3.821e-02  7.624e-03  -5.011 5.41e-07 ***
Average_Change_Rate_Pulse  2.853e-01  1.232e+00   0.232 0.816921
Age                        2.153e-02  7.482e-03   2.877 0.004012 **
Average_Change_Rate_EF    -2.230e+00  8.057e-01  -2.768 0.005644 **
time_since_diagnosis       5.623e-02  9.107e-03   6.174 6.65e-10 ***
Total_No_OP_Visits        -2.205e-03  1.853e-02  -0.119 0.905314
Metoprolol_1_or_02        -7.661e-01  2.128e-01  -3.601 0.000317 ***
spironolactone_1_or_02    -3.461e-01  2.907e-01  -1.191 0.233745
furosemide_1_or_02         8.331e-01  2.058e-01   4.048 5.16e-05 ***
bumetanide_1_or_02         1.422e+00  3.215e-01   4.423 9.73e-06 ***
Max_Number_Meds            8.950e-04  4.890e-04   1.830 0.067194 .
hist_pysch_med2           -5.218e-02  1.976e-01  -0.264 0.791705
Max_Number_Visits         -1.184e-01  6.757e-02  -1.752 0.079720 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5-5: Sample Model Output Summary

Next, from all 15 regression runs, we generate a count of the total number of times that each variable appeared as significant, as shown in Table 5-3. The variables that appear as significant in at least 80% of the runs (i.e., 12 runs or more) are then deemed as the most significant. We include these variables in our proposed final model, referred to as ‘Model_Final’ from this point onwards.

Table 5-3: Number of Significant Appearances of Predictor Variables from the 15 Partition Runs (shown in decreasing order)

No.	Predictor	Type (Categorical or Continuous)	R Identifier	Number of Significant Appearances
1	Language	Categorical/Binary	Language	14
2	Time Since Initial Diagnosis	Continuous	time_since_diagnosis	14
3	Administration of Bumetanide	Categorical/Binary	bumetanide_1_or_0	14
4	Total Number of Medications	Continuous	Max_Number_Meds	14
5	Estimated Income	Continuous	Median_Income_by_Zip	14
6	Marital Status	Categorical/Binary	Marital_status	13
7	Lowest Systolic Blood Pressure	Continuous	Lowest_Systolic_BP	13
8	Administration of Metoprolol	Categorical/Binary	Metoprolol_1_or_0	13
9	Gender	Categorical/Binary	Gender	12
10	Age	Continuous	Age	12
11	Administration of Furosemide	Categorical/Binary	furosemide_1_or_0	11
12	Average Percent Change in Ejection Fraction	Continuous	Average_Change_Rate_EF	11
13	Total Number of Outpatient Visits	Continuous	Total_No_OP_Visits	9
14	Total Number of Cardiology Outpatient Visits	Continuous	Max_Number_Visits	7
15	Administration of Spironolactone	Categorical/Binary	spironolactone_1_or_0	3
16	Presence of History of Psychotic Medications	Categorical/Binary	hist_psych_med	1
17	Average Percent Change in Pulse	Continuous	Average_Change_Rate_Pulse	0

5.4.3 Final Model: Results and Discussion

Our final model (Model_Final) is presented in Table 5-4. In particular, the selected variables from Section 5.4.2 are presented, along with their final coefficients (calculated as the average of all respective coefficients from the 15 random partition runs). It also shows the standard deviation and coefficient of variation associated with each of the predictor variable's coefficients, and the minimum and maximum coefficients that were obtained from all 15 runs.

Table 5-4 Final Logistic Regression Model to Predict First-Time Admissions with the 10 Most Significant Predictors

FINAL MODEL (Model_Final)					
Predictor	**Average β**	**STD**	**C.V.**	**Min Reported β**	**Max Reported β**
Intercept	-1.462	0.343	-23%	-2.294	-0.852
Gender	0.248	0.068	27%	0.141	0.344
Language	0.558	0.106	19%	0.360	0.712
Marital Status	0.387	0.077	20%	0.249	0.516
Estimated Income	-5.750E-06	1.641E-06	-29%	-8.787E-06	-3.441E-06
Lowest Systolic Blood Pressure	-0.011	0.002	-19%	-0.014	-0.006
Age	0.027	0.003	10%	0.022	0.032
Time Since Initial Diagnosis	0.010	0.002	24%	0.005	0.013
Administration of Metoprolol	-0.049	0.043	-88%	-0.128	-0.002
Administration of Bumetanide	1.043	0.145	14%	0.697	1.244
Total Number of Medications	0.003	2.175E-04	8%	0.002	0.003

Before evaluating the final model's performance, we describe the hypothesized intuition behind each of the model's variables. Starting with gender, the final model suggests that males have a higher likelihood of a first-time hospital admission than females. This is consistent with other studies conducted on broad populations of heart failure patients, which associated females

with better heart failure outcomes (e.g., improved survival) [56]. Similarly, the final model suggests that patients that do not speak English have a higher likelihood of a first-time admission than patients that do. We hypothesize that this is due to the studied relationship between limited health literacy and poor communication quality within health care delivery organizations [57]. Marital status is the third significant feature in our model, where single heart failure patients have a greater likelihood of a first-time hospital admission. This is aligned with the findings of a 2009 study which found that heart failure patients with a spouse had better clinical outcomes than patients without (even in the context of depressive symptoms). As a result, the study's recommended intervention was to identify and reinforce alternative social support networks for non-married patients [58]. Estimated income (by median income level of patient's zip code) is the fourth non-clinical feature of our model. The model suggests that patients with lower income levels have a greater likelihood of a first-time admission. We link that to the reported findings on the influence of socioeconomic factors on patient outcomes [59]. Age is the final non-clinical feature that is included in our model. In terms of age, older patients have a greater likelihood of a first-time admission. We hypothesize that this is related to the progressive nature of heart failure and the complex comorbidities that elderly patients more often present with than younger ones.

We now turn to the final model's clinical features, starting with a patient's lowest recorded systolic blood pressure[35] (SBP). The model suggests that the lower the patient's SBP, the greater the likelihood of a first time hospital admission. To explain this finding, we refer to a study reported in the *Journal of the American Medical Association*, where Dr. Gheorghiade and colleagues discuss similar findings by suggesting that "SBP may indicate different stages or

[35] We consider the lowest systolic blood pressure since January 2011, which represents the starting point of all of our analyses' time frames.

pathophysiology of heart failure which accounts for worse outcomes in heart failure patients with lower SBP" [60]. The second clinically-relevant feature in our final model is "time since initial diagnosis". In this case, our model confirms the hypothesis that "*the longer the time elapsed since a patient's first heart failure diagnosis*[36]*, the greater the likelihood of a first-time admission*". This matches the progressive nature of heart failure. In addition to the time elapsed since the initial heart failure diagnosis, the administration of two heart failure specific medications appear as significant model variables. First, the administration of metoprolol appears as significant and has an inverse relationship with likelihood of a first-time hospital admission, i.e., a patient that was administered metoprolol in year $t - 1$ has a lower likelihood of a first-time hospital admission in year t. In contrast, the administration of bumetanide appears as significant and has a direct relationship with likelihood of a first-time admission. In both cases, the findings are in agreement with the hypotheses originally held by the team's cardiologists (discussed earlier in Section 5.2). Finally, the total number of medications administered (in the year prior to admission prediction) appears as the last significant clinical variable, with a positive relationship with likelihood of first-time admission. The intuition here is that a greater number of medications likely points to more comorbidities and, thus, a greater degree of patient clinical complexity. This higher complexity then translates to a greater risk of admission.

As an example, data from a de-identified patient is used, together with this model, to compare our model's predicted outcome of first-time admission with the actual outcome. This is shown in detail in Table 5-5. We notice that this particular patient has the following characteristics: male; 41.6 years old; English not spoken as a primary language; single; estimated annual income of $81,143 (based on median incomes by zip codes); lowest reported systolic

[36] Initial heart failure diagnosis refers to the first reported heart failure diagnosis since January 2011

blood pressure of 124 mmHg; 37.8 months since initial heart failure diagnosis; has seen no administration of metoprolol; has seen an administration of bumetanide; and has a total of 27 medications on record. With such a profile, the patient's heart failure admission outcome is most likely attributed to the significance of (i) social factors like language and marital status, (ii) administration of loop diuretics and (iii) presence of comorbidities (estimated through the total number of medications on record). For this specific male, the other remaining factors are of the hypothesized 'good' nature.

Table 5-5: Sample Run of the Final Model using Data from a De-Identified Heart Failure Patient

(note: β values are the same as those reported in Table 5-4)

EXAMPLE		
Patient Data (x_i)	**Predictor**	**Model Result ($\beta_i x_i$)**
	Intercept (β_0)	1.46E+00
2	Gender (1 = Female; 2 = Male)	4.96E-01
2	Language (1= English; 2 = Not English)	1.12E+00
2	Marital Status (1 = Not Single; 2 = Single)	7.73E-01
81,143	Estimated Income	-4.67E-01
124	Lowest Systolic Blood Pressure	-1.39E+00
41.6	Age	1.12E+00
37.8	Time Since Initial Diagnosis	3.74E-01
1	Administration of Metoprolol (1 = No; 2 = Yes)	-4.90E-02
2	Administration of Bumetanide (1 = No; 2 = Yes)	2.09E+00
27	Total Number of Medications	7.25E-02

SUM = $\beta_0 + \Sigma (\beta_i x_i)$	5.589512143
-1*SUM	-5.589512143
exp(-1*SUM)	0.00373685
1 + exp(-1*SUM)	1.00373685
Estimated Probability = 1/[1 + exp(-1*SUM)]	0.996277062

Threshold	0.5

Predicted Outcome	2	Note: 1 = No Admission; 2 = Admission
Actual Outcome	2	(Notice the effect of the selected threshold)

5.4.4 Final Model: Evaluation of Performance

There are many ways to evaluate the quality of the proposed model. For the purpose of this project, we use the are under the receiver operating curve (AUC). The receiver operating curve (ROC) is useful because it captures all possible thresholds and calculates a true positive rate (sensitivity) and a false positive rate (1-specificity) for each. In other words, it captures the proportion of first-time admissions caught as a function of the proportion of non-first-time admissions that were labeled as first-time admissions. By generating a plot of all thresholds simultaneously and calculating the area under the curve, the overall interpretation we get is the following: "given a random positive and negative (for first-time admissions), what is the proportion of time that we are able to guess which is which correctly?"

Table 5-6 presents the average in-sample and out-of-sample AUCs[37] achieved from all 15 logistic regression runs, and compares them against those of the regression trees and random forests. Interestingly, regression trees and random forests, despite having higher in-sample performances, show a significant drop when tested out of sample. This likely points to an issue of model overfitting when these techniques are applied, and justifies the selection of a final model that is based on logistic regression.

[37] In-sample and out-of-sample AUCs are recorded after each regression run on the 15 training/test partitions. Their averages are then taken as estimates for Model_Final's AUCs.

Table 5-6: Average In-Sample and Out-of-Sample AUCs from Logistic Regression, Regression Tree and Random Forest Modeling Techniques[38]

	Logistic Regression	**Regression Trees**	**Random Forests**
Average In-Sample AUC	0.735	0.737	0.976
Average Out-of-Sample AUC	0.712	0.610	0.655

In addition to average AUCs, we also present the granular in-sample and out-of-sample AUCs achieved by all partition runs, in Figure 5-6 to Figure *5-11*. In each case, we also show the average AUC, the standard deviation (STD), and the upper and lower limits (defined here as (average AUC + STD) and (average AUC – STD), respectively). Coefficient of variation (C.V.) is also reported for all cases. We notice that the out-of-sample C.V. is consistently higher than the in-sample C.V. This is expected due to the greater uncertainty and variability in out-of-sample testing.

[38] Note that the random forest's near-perfect in-sample AUC indicates that model overfitting has most likely taken place. This is also the case with the reported AUCs in Figure 5-10.

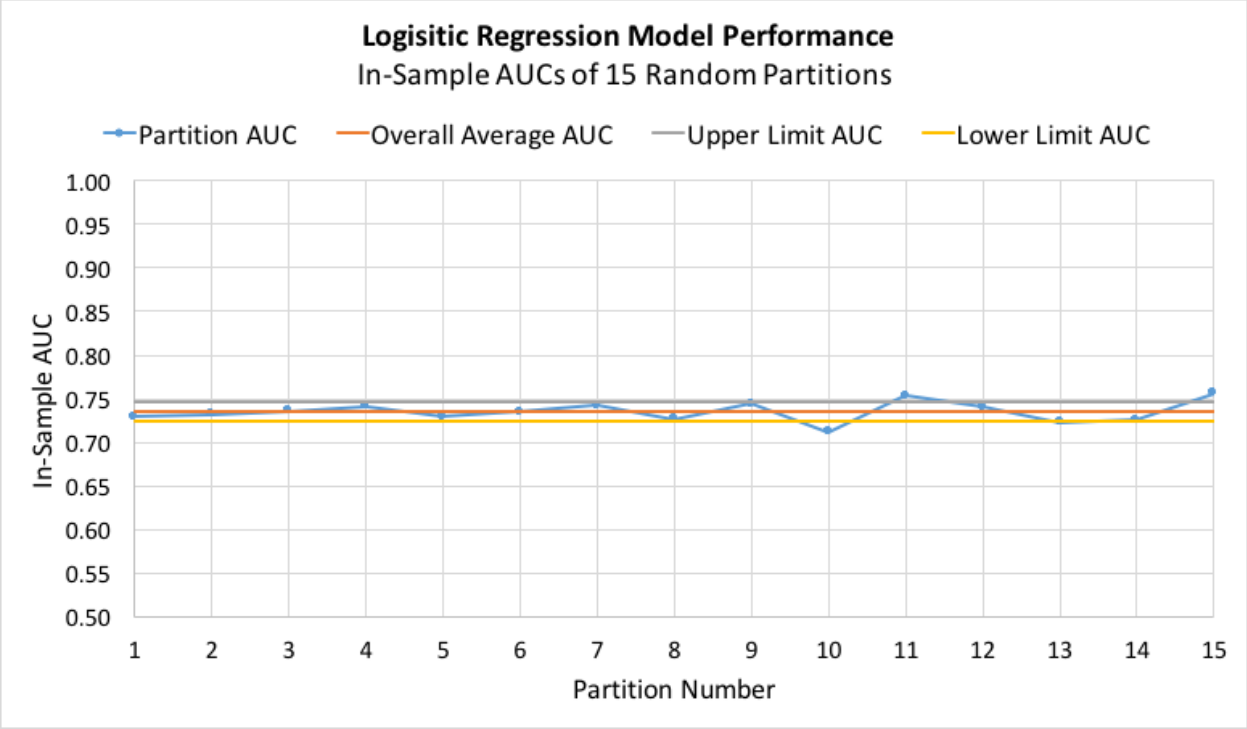


Figure 5-6: Logistic Regression In-Sample AUCs; Average=0.735, STD=0.012, C.V.=1.6%

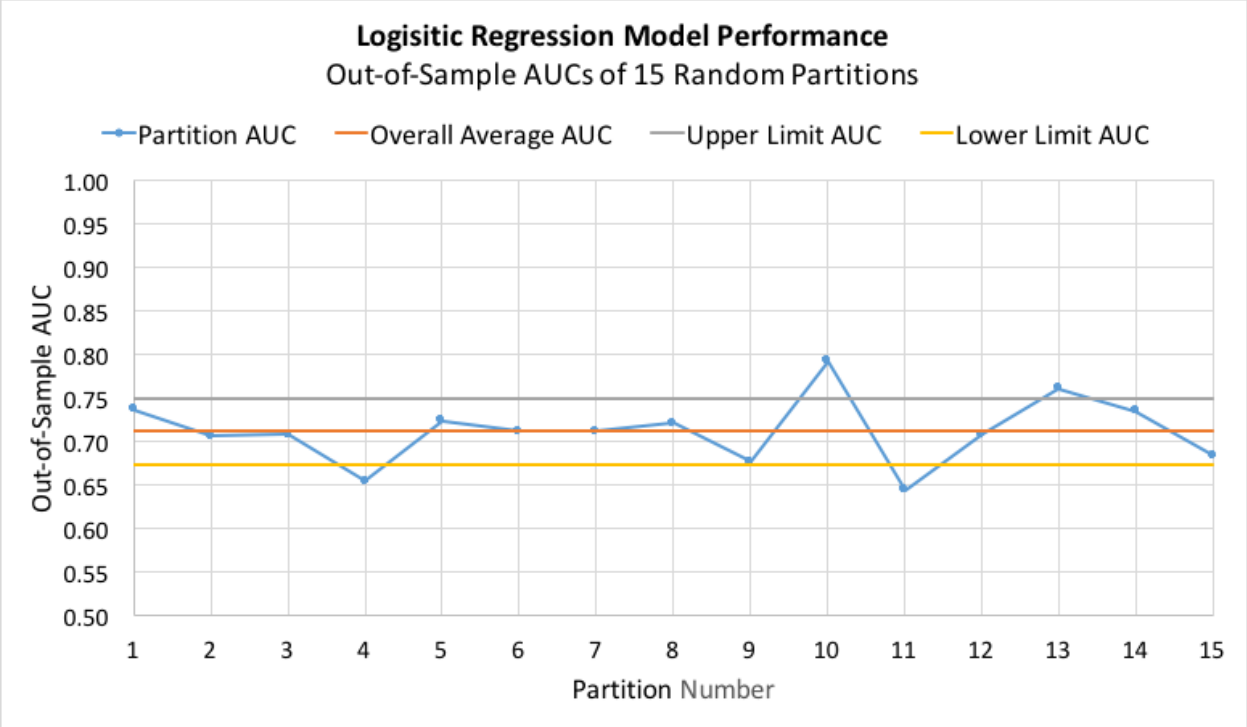


Figure 5-7: Logistic Regression Out-of-Sample AUCs; Average=0.712, STD=0.038, C.V.=5.3%

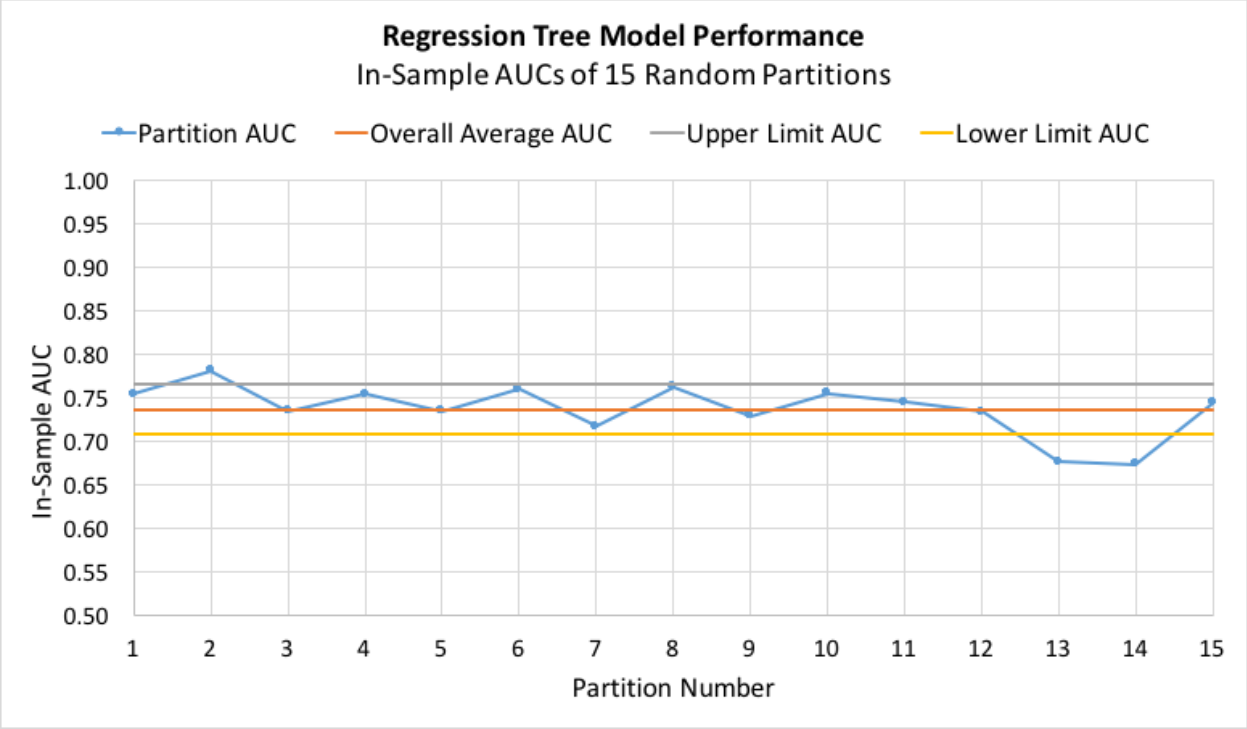


Figure 5-8: Regression Tree In-Sample AUCs; Average=0.737, STD=0.030, C.V.=4.0%

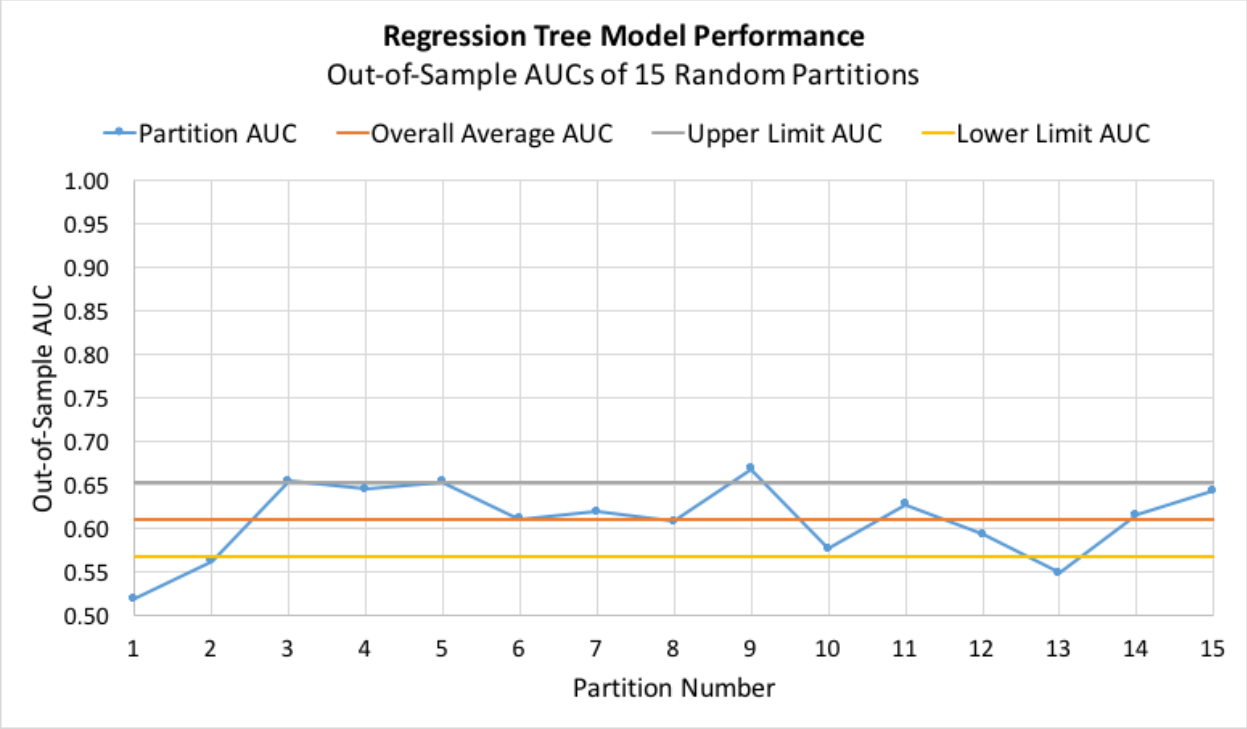


Figure 5-9: Regression Tree Out-of-Sample AUCs; Average=0.610, STD=0.043, C.V.=7.0%

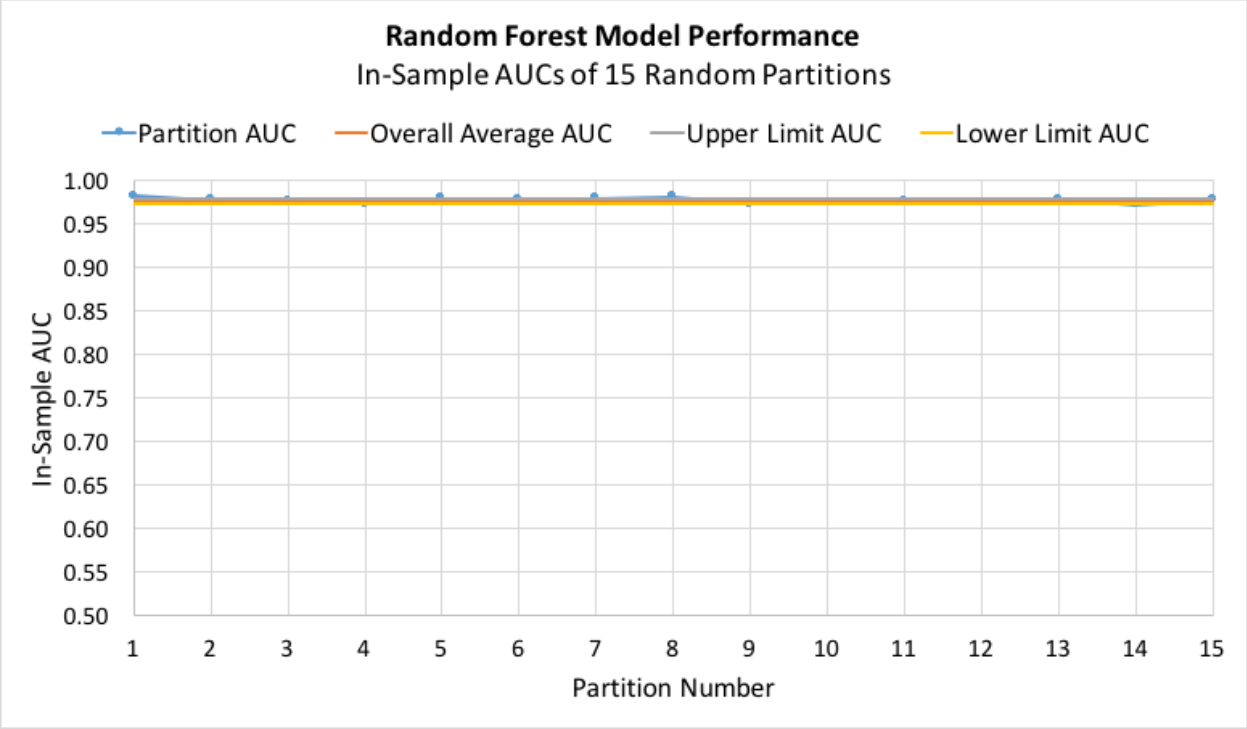


Figure 5-10: Random Forest In-Sample AUCs; Average=0.976, STD=0.003, C.V.=0.3%

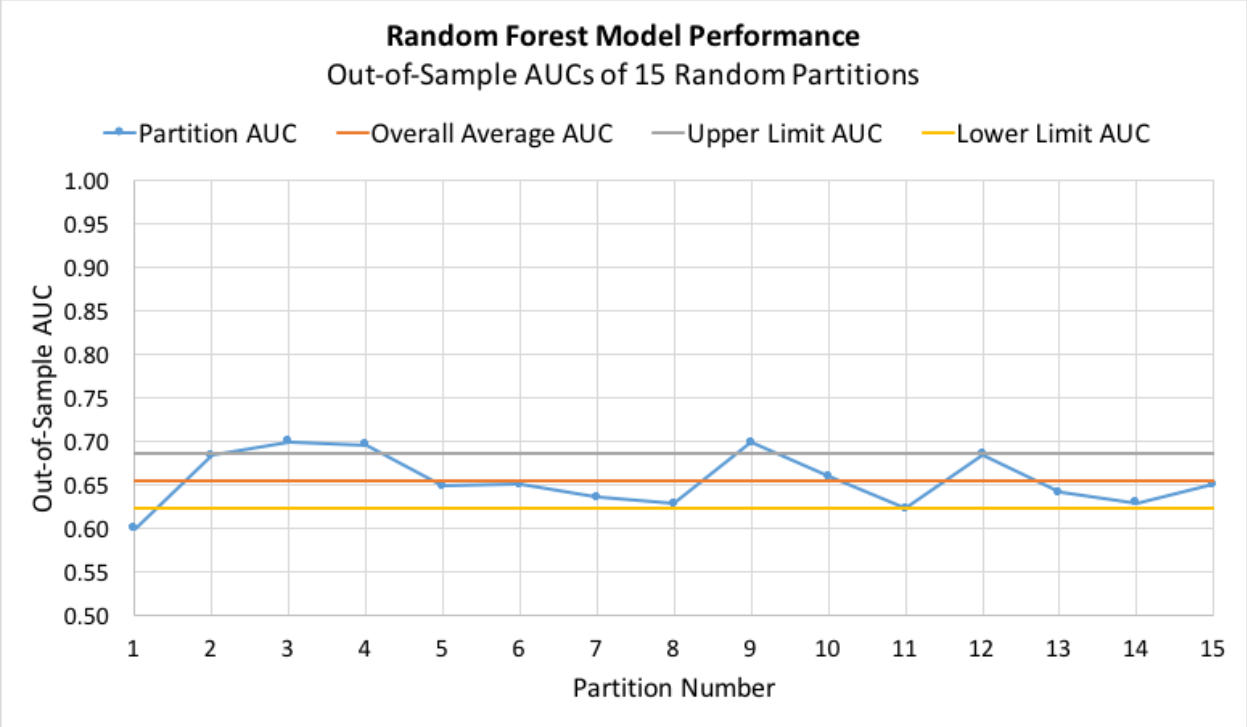


Figure 5-11: Random Forest Out-of-Sample AUCs; Average=0.654, STD=0.023, C.V.=3.6%

5.4.5 Comparison with the Seattle Heart Failure Model

Perhaps one of the most commonly used models in predicting heart failure risk is the Seattle Heart Failure Model (SHFM). Using a cohort of 1,125 heart failure patients, SHFM was developed as a multivariate Cox model to predict 1-, 2- and 3-year survival. As predictors, it uses clinical, pharmacological, device and laboratory features [50].

To validate the performance of our final model against that of more commonly used heart failure risk models, we first consider the AUCs of both models. SHFM has an overall AUC of 0.729 while our final model has a slightly lower overall average AUC of 0.712. However, we believe that the difference of 0.012 in AUC and, consequently, predictive abilities is compensated for by our model's use of simple explanatory variables.

In terms of predictors, SHFM highly depends on the results of clinical, laboratory and pharmacological tests. This constitutes a list of complex data entries that are often hard to obtain and input by patients themselves. For example, to obtain a risk score, SHFM requires the input of levels of sodium, total cholesterol and hemoglobin, to name a few [50]. To model risk scores appropriately, such features require dedicated blood tests, accurate lab measurements and proper reporting of results. As such, it is not unusual that using SHFM and interpreting its outputs often requires the assistance of a healthcare provider. In fact, the first question that users are asked when attempting to calculate a risk score using an online SHFM calculator is whether or not they are healthcare providers. If the answer is 'No', users are informed that "the calculator is designed to be used by healthcare providers" and are encouraged "to print a copy of the calculator and take it to their healthcare providers" [61].

Our model, on the other hand, requires the input of very simple variables, most of which can actually be reported by patients themselves. Systolic blood pressure is the only clinical

indicator of significance in our proposed model, and it can usually be measured at home with the help of a simple blood pressure monitor. The only variables we foresee as potentially posing challenges for patients attempting to calculate their risk scores are time since initial diagnosis, presence of history of metoprolol administration, presence of history of bumetanide administration, and total number of medications. Even then, these can be easily obtained during scheduled outpatient follow-ups and medication reconciliations or through a phone call with the provider.

5.5 *Potential for Future Research and Model Improvement*

As it currently stands, the value of our final proposed model lies in its ability to predict heart failure risk (on the basis of probability of admission for heart failure) with relatively simple features. In terms of future research, several steps are suggested here for their potential to further improve our model's predictive power.

5.5.1 Incorporating Additional Predictive Variables

First, the incorporation of medication doses and frequencies is proposed. Currently, heart failure medications included in the model are limited to a binary variable on whether or not the medication was administered at least once, which limits the risk scoring to just the 'directionality' of the medications' effects. Extending this variable to include dosages would permit the testing of more meaningful hypotheses such as the "beneficial effect of increased dosages of beta-blockers on reducing heart failure admission outcomes". However, to facilitate an appropriate and meaningful inclusion of dosages, medication doses and frequencies would need to be collected in a structured manner for all patients. As it stands now, medications data can be sourced from multiple locations (refer to data sources discussed in Section 3.2) and only

one of these sources, specifically LMR, provides consistent data on doses and frequencies[39]. Unfortunately, only a minority of medications data is currently available and sourced from LMR (~37.6% on average). As such, for an effective inclusion of this data in the future, including complete and consistent medications data, in LMR format, as part of the heart failure registry proposed in Section 6.1 is recommended.

Lab test results constitute a second feature that is proposed for future inclusion in our predictive model. In terms of first lab tests proposed for inclusion, we recommend incorporating N-terminal pro b-type natriuretic peptide (NT-proBNP) and blood urea nitrogen (BUN)[40] as a first step. Next, test results included in other heart failure risk models could be explored. For example, similar to the Seattle Heart Failure Model, the inclusion of hemoglobin, serum sodium, lymphocytes or uric acid, could prove beneficial in the improvement of predictive power. However, awareness should be made that this comes at the expense of potentially increased model complexity.

The third and fourth proposed predictors for inclusion are the New York Heart Association (NYHA) functional classes and the American College of Cardiology/American Heart Association (ACC/AHA) heart failure classification stages, respectively. The main challenge with including NYHA classes and ACC/AHA stages is that they are not documented for all

[39] This is due to the nature of LMR, a 'home grown' classification system developed at MGH, which performs a number of analytic processes in order to make the data as consistent and reliable as possible for the data systems it feeds into (i.e., RPDR) and users.

[40] NT-proBNP and BUN are proposed as the first lab tests due to the availability of clinical direction from our team's cardiologists on cutoffs that are typically associated with 'worse' heart failure admission outcomes. Specifically, NT-proBNP levels higher than 1,000 pg/mL and BUN levels higher than 20 mg/dL are hypothesized to result in higher risk scores.

patients and are often available only for patients managed by the heart failure team. Second, where available, this data is documented as free text in patient notes and would therefore require the development of an appropriate text search routine for extraction. Finally, as mentioned earlier, increased model complexity is another factor to consider when including additional predictor variables.

5.5.2 Classifying Heart Failure Patients and Developing 'More Focused' Models

In addition to the incorporation of more predictive variables, the classification of heart failure patients and development of 'more focused' models for different classes is suggested as another way of potentially achieving greater predictive power. As it stands now, our model does not classify for any specific groups and treats all heart failure equally.

One suggested method of classification is by left ventricular ejection fraction (LVEF). In developing our model, LVEF was included as one of the 27 predictor variables. However, throughout the process of model training with multiple iterations, it did not appear as a statistically significant predictor of first-time admissions and, hence, is not included in the final model. Perhaps a more useful way of using LVEF is to group patients into the following two categories: (i) heart failure with reduced ejection fraction (HFrEF) where LVEF is below 40% and (ii) heart failure with preserved ejection fraction (HFpEF) where LVEF is above 50%. Using the same predictors, we can then train the model on HFrEF patients to test whether or not an improvement in out-of-sample AUC is observed. The hypothesis that the model would show greater relevance and better predictive power with HFrEF patients derives from the fact that the medications used as part of our model's predictors are actually associated with HFrEF patients. In other words, according to clinical guidance, higher dosages of beta blockers, ACE inhibitors, ARBs and MRAs are typically associated with better hospital admission outcomes for HFrEF

patients specifically. Similarly, higher dosages of loop diuretics are often associated with worse outcomes. No similar recommendations on medications currently exist for HFpEF patients.

Alternative ways of grouping patients before modeling include classification by NYHA functional class, ACC/AHA stage, or demographics. Irrespective of the method of classification, training separate models and exploring alternative modeling techniques such as decision tree modeling and random forests are recommended for the different classes.

5.5.3 Addressing Variable Non-Linearities

As mentioned previously, our model boasts simplicity as a feature, part of which is modeling the effects of all predictors linearly as first-order terms. Clearly, a limitation of this approach is that variable non-linearities are not captured.

With regards to addressing non-linearities, we propose exploring the modeling of variables, x_i, in various transformed versions (e.g., $\log(x_i)$, x_i^2, $\sqrt{x_i}$, etc.). As an example, age does not always follow a linear relationship with outcomes and it is not unusual to see behavior where good outcomes are observed within a specific age range and worse ones outside. In such cases, modeling age as a straight line can pose limitations to the model's predictive power. The best way to deal with non-linearities is to start by plotting the independent variable against the target variable being predicted, and to observe its behavior visually. Once again, we caution that this comes at the expense of increased complexity and, possibly, model overfitting.

5.5.4 Expanding the Model to Predict Overall Admissions

Finally, we propose expanding the model to predict overall admissions. This would essentially be the same exercise, with the slight modification being that we now consider all patients linked to an MGH primary care provider, as opposed to just those with no prior admission history.

Similarly, we train the model on any type of heart failure admission outcome and do not limit it to first-time heart failure admissions.

Moreover, we propose predicting for all-cause hospitalization and heart failure hospitalization separately. For the latter, we also propose parsing the hospitalizations into those with heart failure as their primary discharge diagnosis versus those with heart failure as their secondary, or tertiary, discharge diagnosis.

6. Recommendations and Conclusions

6.1 *Operational Recommendations*

The final model presented in Section 5.4.3 shows that some of the most significant predictors of heart failure hospitalization patterns are non-physiologic features. This is expected of chronic heart failure, a condition that is heavily influenced by individual characteristics and environmental factors reflecting a patient's lifestyle.

To provide better care systems for heart failure patients that would, in turn, translate to lower hospitalization rates and generate hospital capacity for other inpatient cases, we propose taking some initial steps by developing a registry of heart failure patients within cardiology and primary care clinics that tracks the features listed in Table 5-5 (and perhaps others) as well as the appointments scheduled with any outpatient clinic. This registry would then prove useful in assigning patients a risk score, based on our proposed model in Table 5-4, that can be used for allocating a range of tailored interventions and care resources effectively. In addition, we propose building clear communication paths between primary care, cardiology and the hospital, as well as creating a system that can routinely take care of heart failure patients as opposed to the current transient care structure.

Furthermore, in terms of interventions, we propose to start with addressing the socioeconomic features that are predicted to have a greater effect on heart failure admissions. While some factors are more challenging to address than others (e.g., gender and marital status), programs can be designed to address, say, the factor of English not being spoken as a primary language. One example is introducing interpreter-based outpatient education that aims to ensure periodic communication with high-risk patients and promote the importance of proper heart

failure management. By assessing and confirming a patient's language needs and preferences, the hope is that patients would be better equipped with the skills required to recognize escalating symptoms, follow activity and exercise recommendations, adhere to prescribed medication instructions and diet guidelines, and understand the rationale and importance of follow-up appointments, not to count the expected increase in patient satisfaction as well.

Moreover, we propose using the final model's features and risk scores to better allocate remote monitoring based interventional resources (e.g., telemonitoring) for the appropriate patients. In particular, patients currently qualify for enrollment in the *Connected Cardiac Care Program* based on the criteria listed in Section 2.3.1. A closer look at the last three points shows that there is potentially room for improvement in the assignment guidelines of telemonitoring. More specifically, it might be worthwhile to assign the remote monitoring to patients that are unable to communicate in English as opposed to those that do (as per the current requirement). Similarly, since marital status was observed to have a great predictive effect on heart failure admissions (likely indicative of the importance of a family or support system for the patient), we propose lifting the constraint of assigning telemonitoring only to patients that have a caregiver at their disposal. Finally, the requirement of a clean and equipped home setting is likely misaligned with the significant feature of a patient's median income by zip code. As such, reevaluation of this criterion is recommended.

On a more long-term horizon, we propose incorporating more advanced and continuous monitoring techniques as part of the heart failure care management processes at MGH. For example, as more patients with implantable hemodynamic monitors and sensors become part of the MGH and Partners network, we propose making use of the valuable data collected by these devices to enable better monitoring of patients and near 'real-time' adjustments as needed. The

premise would be that continuous measurements of intracardiac filling pressures provide a good indicator of a patient's volume status which, in turn, could give providers more direction on, say, the necessary adjustments in diuretic therapy.

6.2 Conclusions

This projected was charged with the following goal: develop new systems and processes that will provide better care for heart failure patients and more control for their conditions. Through an exhaustive mapping of the heart failure care pathway, a rigorous analysis of the granularities of heart failure patient trends, and the implementation of predictive modeling techniques, we develop a risk scoring mechanism that identifies patients with the greatest likelihood of a first-time heart failure admission. Along with other system-level operational recommendations, we propose this model, the value of which lies in its simplicity and use of easily obtained variables, as a prioritization technique to guide the allocation of limited outpatient resources to high-risk patients. The aim is that this would foster a proactive nature of outpatient monitoring and care which, among other things, would then translate into a reduction or elimination of avoidable heart failure admissions.

Appendix A

Below is a brief description of various types of home services that were discussed in Section 3.2:

Type of Discharge Services	Description
Visiting Nurse Association (VNA)	The VNA is an association that supports, promotes and advocates for providers of home health, hospice and palliative care. VNA members provide cost-effective and compassionate home health and hospice care to the most vulnerable patients, particularly the elderly and individuals with disabilities.
Partners Mobile Observation Unit (PMOU)	The PMOU program is an initiative created by Partners HealthCare at Home (PHH) and Partners Population Health Management (PHM) to help reduce hospitalizations. The program involves a team that works closely with physicians in the ED and Observation Unit (OBS), as well as the Case Management and Integrated Care Management departments. PMOU aims to bring a clinicians with advanced skills into the homes of patients to directly assess their needs.
Rehabilitation Services	These include services like physical, occupational, speech-language therapy and cognitive therapy, where needed. For example, PHH physical therapists help restore function, improve patient mobility, and relieve pain or prevent/limit disability.
Skilled Nursing Facility (SNF)	SNFs offer short-term care for patients recovering from an illness or injury, as well as long-term care for those who no longer can stay at home. A wide range of rehabilitation services are offered at SNFs and are customized to meet each patient's needs.

Appendix B

The following ICD-9 codes were used to identify heart failure principal diagnoses:

402.01	404.13	428.21	428.33
402.11	404.91	428.22	428.40
402.91	404.93	428.23	428.41
404.01	428.0	428.30	428.42
404.03	428.1	428.31	428.43
404.11	428.20	428.32	428.9

The following ICD-10 codes were used to identify heart failure principal diagnoses:

I10	I15.8	I50.21	I50.43	I87.391
I11.0	I15.9	I50.22	I50.9	I87.392
I11.9	I20.0	I50.23	I51.2	I87.393
I12.0	I25.2	I50.30	I51.5	I87.399
I12.9	I25.3	I50.31	I51.7	I95.89
I13.0	I31.2	I50.32	I51.81	I97.0
I13.2	I31.4	I50.33	I87.301	I97.110
I15.0	I42.8	I50.40	I87.302	I97.190
I15.1	I50.1	I50.41	I87.303	I97.191
I15.2	I50.20	I50.42	I87.309	I97.3

Appendix C

Below is a brief description of the clinical indicators and features used in our predictive models:

Clinical Indicator	Description
Systolic Blood Pressure (BP)	The blood pressure when the heart is contracting. It is specifically the maximum arterial pressure during contraction of the left ventricle of the heart. The time at which ventricular contraction occurs is called systole. In a blood pressure reading, the systolic pressure is typically the first number recorded. For example, with a blood pressure of 120/80 ("120 over 80"), the systolic pressure is 120.
Pulse	Heart rate, or pulse, is the number of times the heart beats per minute.
Ejection Fraction	Ejection fraction is a measurement of the percentage of blood leaving the heart each time it contracts with a heartbeat. According to the Mayo Clinic, an LV ejection fraction of 55% or higher is considered normal. An LV ejection fraction of 50% or lower is considered reduced. Experts vary in their opinions about an ejection fraction between 50% and 55%.

Appendix D

The following represents a list of medications used to identify patients with a history of psychotic conditions:

1. **Anti-Depressants:**
    - Selective Serotonin Reuptake Inhibitor (SSRI's):
        - Prozac
        - Zoloft
        - Paxil
        - Celexa
        - Lexapro
    - Serotonin–Norepinephrine Reuptake Inhibitor (SNRI's):
        - Effexor
        - Pristiq
        - Cymbalta
        - Savella
    - Monoamine Oxidase Inhibitors (MAOI's):
        - Parnate
        - Nardil
        - Deprenyl
    - Others:
        - Wellbutrin
        - Remeron
2. **Anti-Psychotics:**
    - Typical or First Generation:
        - Haldol
    - Atypical or Second Generation:
        - Abilify
        - Clozaril
        - Fanapt
        - Zyprexa
        - Invega
        - Seroquel
        - Risperdal
        - Geodon
3. **Dementia Medications:**
    - Cholinesterase Inhibitors:
        - Aricept
        - Exelon
        - Razadyne
        - Cognex
    - N-Methyl-D-Aspartate (NMDA) Antagonists:
        - Namenda

Appendix E

The following ICD-9 codes were used to identify patients with a history of substance abuse:

1. Alcohol Use/Abuse:

303	303.9
303.01	303.91
303.02	303.92
303.03	303.93

2. Polysubstance Abuse:

304	304.22	304.5	304.72
304.01	304.23	304.51	304.73
304.02	304.3	304.52	304.8
304.03	304.31	304.53	304.81
304.1	304.32	304.6	304.82
304.11	304.33	304.61	304.83
304.12	304.4	304.62	304.9
304.13	304.41	304.63	304.91
304.2	304.42	304.7	304.92
304.21	304.43	304.71	304.93

Appendix F

No.	Predictor	R Identifier	Type (Categorical or Continuous)	Hypothesized Relationship with HF Admission Outcome	Category
1	Gender	Gender	Categorical/Binary	+ (Male)	**Patient Characteristics**
2	Language	Language	Categorical/Binary	+ (Not English)	
3	Age	Age	Continuous	+	
4	Marital Status	Marital_status	Categorical/Binary	+ (Single)	
5	Administration of Metoprolol	Metoprolol_1_or_0	Categorical/Binary	– ('1')	**HF-Related Medications (beta-blockers, ARB, MRAs, loop diuretics)**
6	Administration of Carvedilol	carvedilol_1_or_0	Categorical/Binary	– ('1')	
7	Administration of Spironolactone	spironolactone_1_or_0	Categorical/Binary	– ('1')	
8	Administration of Valsartan	valsartan_1_or_0	Categorical/Binary	– ('1')	
9	Administration of Furosemide	furosemide_1_or_0	Categorical/Binary	+ ('1')	
10	Administration of Torsemide	torsemide_1_or_0	Categorical/Binary	+ ('1')	
11	Administration of Bumetanide	bumetanide_1_or_0	Categorical/Binary	+ ('1')	
12-14	Systolic BP (lowest / mean / avg change)	Lowest_Systolic_BP / Mean_Systolic_BP / Average_Change_Rate_Systolic_BP	Continuous	–	**Clinical Indicators**
15-17	Pulse (highest / mean / avg change)	Max_Pulse / Mean_Pulse / Average_Change_Rate_Pulse	Continuous	+	
18-20	Ejection Fraction (minimum / mean / avg change)	Min_EF / Mean_EF / Average_Change_Rate_EF	Continuous	–	
21	Total Number of Cardiology Outpatient Visits in Previous Year	Max_Number_Visits	Continuous	–/+	**Institutional Ownership and Engagement**
22	Total Number of Outpatient Visits	Total_No_OP_Visits	Continuous	–/+	
23	Time Since Initial Diagnosis	time_since_diagnosis	Continuous	+	
24	Presence of History of Substance Abuse	substance_abuse_1_or_0	Categorical/Binary	+ ('1')	**Others (socio-economic factors, comorbidities, etc.)**
25	Presence of History of Psychotic Medications	hist_psych_med	Categorical/Binary	+ ('1')	
26	Total Number of Medications	Max_Number_Meds	Continuous	+	
27	Zip Code Median Income (as 'Estimated Income')	Median_Income_by_Zip	Continuous	–	

Appendix G

Result summaries of the 15 partition regression runs discussed in Section 5.4.2 are presented here:

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5111  -1.0789  -0.1546   1.1150   1.7004

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          -8.519e-01  2.945e-01  -2.893 0.003820 **
Gender2               1.409e-01  6.086e-02   2.315 0.020620 *
Language2             3.597e-01  9.513e-02   3.781 0.000156 ***
Marital_status2       2.871e-01  5.993e-02   4.792 1.65e-06 ***
Median_Income_by_Zip -7.798e-06  1.132e-06  -6.887 5.71e-12 ***
Lowest_Systolic_BP   -1.201e-02  1.908e-03  -6.295 3.08e-10 ***
Age                   2.548e-02  2.279e-03  11.182  < 2e-16 ***
time_since_diagnosis  6.789e-03  2.342e-03   2.899 0.003747 **
Metoprolol_1_or_02   -6.660e-02  6.791e-02  -0.981 0.326773
bumetanide_1_or_02    1.052e+00  1.343e-01   7.836 4.65e-15 ***
Max_Number_Meds       2.368e-03  2.230e-04  10.619  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6983.3  on 5437  degrees of freedom
AIC: 7005.3

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8192  -1.0586  -0.1451   1.1147   1.8088

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.459e+00  2.990e-01  -4.881 1.06e-06 ***
Gender2               3.213e-01  6.180e-02   5.198 2.01e-07 ***
Language2             4.653e-01  9.626e-02   4.833 1.34e-06 ***
Marital_status2       3.442e-01  6.045e-02   5.694 1.24e-08 ***
Median_Income_by_Zip -5.452e-06  1.116e-06  -4.884 1.04e-06 ***
Lowest_Systolic_BP   -1.235e-02  1.964e-03  -6.291 3.15e-10 ***
Age                   2.756e-02  2.403e-03  11.468  < 2e-16 ***
time_since_diagnosis  1.057e-02  2.383e-03   4.437 9.11e-06 ***
Metoprolol_1_or_02   -4.654e-03  6.803e-02  -0.068    0.945
bumetanide_1_or_02    1.199e+00  1.335e-01   8.982  < 2e-16 ***
Max_Number_Meds       2.620e-03  2.232e-04  11.737  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6880.9  on 5437  degrees of freedom
AIC: 6902.9

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8892  -1.0616  -0.1477   1.1083   1.7792

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -1.881e+00  2.963e-01  -6.349 2.17e-10 ***
Gender2              2.346e-01  6.211e-02   3.777 0.000158 ***
Language2            6.421e-01  9.454e-02   6.791 1.11e-11 ***
Marital_status2      4.730e-01  6.062e-02   7.803 6.04e-15 ***
Median_Income_by_Zip -3.622e-06  1.121e-06  -3.230 0.001237 **
Lowest_Systolic_BP   -9.823e-03  1.930e-03  -5.088 3.61e-07 ***
Age                  2.831e-02  2.317e-03  12.220  < 2e-16 ***
time_since_diagnosis 9.073e-03  2.328e-03   3.897 9.75e-05 ***
Metoprolol_1_or_02   -2.793e-02  6.856e-02  -0.407 0.683742
bumetanide_1_or_02    9.975e-01  1.355e-01   7.363 1.79e-13 ***
Max_Number_Meds       2.592e-03  2.305e-04  11.246  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6898.4  on 5437  degrees of freedom
AIC: 6920.4

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0396  -1.0500  -0.1842   1.1263   1.7259

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -1.502e+00  2.940e-01  -5.110 3.22e-07 ***
Gender2              1.462e-01  6.146e-02   2.378   0.0174 *
Language2            6.748e-01  9.420e-02   7.164 7.86e-13 ***
Marital_status2      2.492e-01  6.043e-02   4.124 3.72e-05 ***
Median_Income_by_Zip -4.450e-06  1.107e-06  -4.020 5.81e-05 ***
Lowest_Systolic_BP   -9.540e-03  1.924e-03  -4.960 7.05e-07 ***
Age                  2.475e-02  2.315e-03  10.691  < 2e-16 ***
time_since_diagnosis 9.373e-03  2.357e-03   3.977 6.98e-05 ***
Metoprolol_1_or_02   -1.765e-02  6.714e-02  -0.263   0.7927
bumetanide_1_or_02    1.047e+00  1.295e-01   8.085 6.21e-16 ***
Max_Number_Meds       2.913e-03  2.263e-04  12.874  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6884.3  on 5437  degrees of freedom
AIC: 6906.3

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7451  -1.0428  -0.1361   1.0964   1.8177

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -1.395e+00  3.055e-01  -4.566 4.97e-06 ***
Gender2              2.191e-01  6.264e-02   3.498 0.000469 ***
Language2            6.417e-01  9.342e-02   6.869 6.46e-12 ***
Marital_status2      3.841e-01  6.098e-02   6.298 3.01e-10 ***
Median_Income_by_Zip -8.787e-06  1.171e-06  -7.503 6.24e-14 ***
Lowest_Systolic_BP   -1.113e-02  1.949e-03  -5.710 1.13e-08 ***
Age                  2.782e-02  2.380e-03  11.692  < 2e-16 ***
time_since_diagnosis  1.153e-02  2.352e-03   4.904 9.38e-07 ***
Metoprolol_1_or_02   -8.451e-02  6.863e-02  -1.231 0.218172
bumetanide_1_or_02    1.158e+00  1.328e-01   8.723  < 2e-16 ***
Max_Number_Meds       2.829e-03  2.252e-04  12.560  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6769.5  on 5437  degrees of freedom
AIC: 6791.5

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5749  -1.0639  -0.1651   1.1227   1.7898

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -1.278e+00  2.949e-01  -4.334 1.46e-05 ***
Gender2              2.740e-01  6.108e-02   4.485 7.28e-06 ***
Language2            5.474e-01  9.528e-02   5.745 9.20e-09 ***
Marital_status2      4.395e-01  5.994e-02   7.332 2.26e-13 ***
Median_Income_by_Zip -3.564e-06  1.092e-06  -3.264  0.00110 **
Lowest_Systolic_BP   -1.320e-02  1.941e-03  -6.797 1.07e-11 ***
Age                  2.534e-02  2.370e-03  10.693  < 2e-16 ***
time_since_diagnosis  7.080e-03  2.411e-03   2.937  0.00331 **
Metoprolol_1_or_02   -2.116e-03  6.878e-02  -0.031  0.97546
bumetanide_1_or_02    9.321e-01  1.287e-01   7.245 4.32e-13 ***
Max_Number_Meds       2.705e-03  2.260e-04  11.969  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6922.4  on 5437  degrees of freedom
AIC: 6944.4

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9667  -1.0452  -0.1613   1.1224   1.8339

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -1.590e+00  2.983e-01  -5.328 9.90e-08 ***
Gender2              3.387e-01  6.282e-02   5.392 6.97e-08 ***
Language2            5.343e-01  9.596e-02   5.567 2.59e-08 ***
Marital_status2      4.858e-01  6.129e-02   7.927 2.25e-15 ***
Median_Income_by_Zip -3.441e-06  1.120e-06  -3.074  0.00211 **
Lowest_Systolic_BP   -1.230e-02  1.937e-03  -6.347 2.20e-10 ***
Age                  2.553e-02  2.363e-03  10.804  < 2e-16 ***
time_since_diagnosis  1.118e-02  2.364e-03   4.729 2.25e-06 ***
Metoprolol_1_or_02   -6.737e-02  6.886e-02  -0.978  0.32795
bumetanide_1_or_02    1.127e+00  1.310e-01   8.603  < 2e-16 ***
Max_Number_Meds       2.900e-03  2.254e-04  12.867  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6817.6  on 5437  degrees of freedom
AIC: 6839.6

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6360  -1.0613  -0.1437   1.1044   1.8293

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -1.441e+00  3.097e-01  -4.655 3.24e-06 ***
Gender2              2.817e-01  6.300e-02   4.472 7.75e-06 ***
Language2            4.935e-01  9.530e-02   5.179 2.23e-07 ***
Marital_status2      3.146e-01  6.151e-02   5.114 3.15e-07 ***
Median_Income_by_Zip -8.243e-06  1.188e-06  -6.941 3.90e-12 ***
Lowest_Systolic_BP   -1.238e-02  1.953e-03  -6.338 2.33e-10 ***
Age                  3.110e-02  2.384e-03  13.049  < 2e-16 ***
time_since_diagnosis  9.110e-03  2.368e-03   3.848 0.000119 ***
Metoprolol_1_or_02   -1.279e-01  6.920e-02  -1.848 0.064642 .
bumetanide_1_or_02    9.170e-01  1.326e-01   6.913 4.74e-12 ***
Max_Number_Meds       2.882e-03  2.297e-04  12.543  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6858.3  on 5437  degrees of freedom
AIC: 6880.3

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0584  -1.0507  -0.1434   1.1082   1.8388

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)          -2.294e+00  3.070e-01  -7.474 7.79e-14 ***
Gender2               2.536e-01  6.196e-02   4.093 4.26e-05 ***
Language2             6.335e-01  9.535e-02   6.643 3.06e-11 ***
Marital_status2       4.031e-01  6.073e-02   6.637 3.19e-11 ***
Median_Income_by_Zip -6.479e-06  1.144e-06  -5.665 1.47e-08 ***
Lowest_Systolic_BP   -5.951e-03  1.924e-03  -3.093  0.00198 **
Age                   2.867e-02  2.340e-03  12.251  < 2e-16 ***
time_since_diagnosis  1.333e-02  2.379e-03   5.601 2.13e-08 ***
Metoprolol_1_or_02   -8.088e-03  6.718e-02  -0.120  0.90417
bumetanide_1_or_02    1.143e+00  1.355e-01   8.436  < 2e-16 ***
Max_Number_Meds       2.999e-03  2.262e-04  13.257  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6815.7  on 5437  degrees of freedom
AIC: 6837.7

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8063  -1.0636  -0.1394   1.0967   1.8268

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.163e+00  2.938e-01  -3.960 7.48e-05 ***
Gender2               3.442e-01  6.194e-02   5.557 2.75e-08 ***
Language2             5.275e-01  9.413e-02   5.603 2.10e-08 ***
Marital_status2       4.501e-01  6.021e-02   7.475 7.73e-14 ***
Median_Income_by_Zip -5.621e-06  1.120e-06  -5.018 5.22e-07 ***
Lowest_Systolic_BP   -1.404e-02  1.947e-03  -7.214 5.44e-13 ***
Age                   2.659e-02  2.296e-03  11.579  < 2e-16 ***
time_since_diagnosis  8.995e-03  2.360e-03   3.812 0.000138 ***
Metoprolol_1_or_02   -7.333e-02  6.784e-02  -1.081 0.279691
bumetanide_1_or_02    1.025e+00  1.329e-01   7.716 1.20e-14 ***
Max_Number_Meds       2.518e-03  2.323e-04  10.838  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6914.6  on 5437  degrees of freedom
AIC: 6936.6

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7523  -1.0771  -0.1677   1.1307   1.7311

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.386e+00  2.932e-01  -4.727 2.27e-06 ***
Gender2               2.614e-01  6.145e-02   4.255 2.09e-05 ***
Language2             5.268e-01  9.421e-02   5.592 2.25e-08 ***
Marital_status2       3.073e-01  5.970e-02   5.147 2.64e-07 ***
Median_Income_by_Zip -5.455e-06  1.107e-06  -4.929 8.28e-07 ***
Lowest_Systolic_BP   -8.317e-03  1.929e-03  -4.312 1.62e-05 ***
Age                   2.380e-02  2.272e-03  10.474  < 2e-16 ***
time_since_diagnosis  5.128e-03  2.317e-03   2.213   0.0269 *
Metoprolol_1_or_02   -8.844e-03  6.748e-02  -0.131   0.8957
bumetanide_1_or_02    1.111e+00  1.335e-01   8.326  < 2e-16 ***
Max_Number_Meds       2.382e-03  2.166e-04  10.994  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 7010.1  on 5437  degrees of freedom
AIC: 7032.1

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5866  -1.0694  -0.1198   1.0960   1.8073

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.777e+00  3.001e-01  -5.921 3.20e-09 ***
Gender2               1.666e-01  6.172e-02   2.699  0.00695 **
Language2             3.933e-01  9.688e-02   4.060 4.91e-05 ***
Marital_status2       3.694e-01  6.046e-02   6.110 9.97e-10 ***
Median_Income_by_Zip -6.301e-06  1.140e-06  -5.525 3.29e-08 ***
Lowest_Systolic_BP   -1.121e-02  1.926e-03  -5.817 5.98e-09 ***
Age                   3.236e-02  2.366e-03  13.677  < 2e-16 ***
time_since_diagnosis  1.067e-02  2.321e-03   4.596 4.30e-06 ***
Metoprolol_1_or_02   -8.398e-02  6.846e-02  -1.227  0.21994
bumetanide_1_or_02    8.576e-01  1.367e-01   6.274 3.52e-10 ***
Max_Number_Meds       2.802e-03  2.202e-04  12.728  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6907.7  on 5437  degrees of freedom
AIC: 6929.7

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6490  -1.0692  -0.1681   1.1285   1.7205

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.069e+00  2.925e-01  -3.654 0.000258 ***
Gender2               2.080e-01  6.113e-02   3.402 0.000668 ***
Language2             5.331e-01  9.367e-02   5.691 1.26e-08 ***
Marital_status2       4.041e-01  5.952e-02   6.790 1.12e-11 ***
Median_Income_by_Zip -5.961e-06  1.145e-06  -5.206 1.93e-07 ***
Lowest_Systolic_BP   -1.061e-02  1.922e-03  -5.520 3.39e-08 ***
Age                   2.161e-02  2.279e-03   9.485  < 2e-16 ***
time_since_diagnosis  9.437e-03  2.339e-03   4.034 5.49e-05 ***
Metoprolol_1_or_02   -1.415e-02  6.700e-02  -0.211 0.832760
bumetanide_1_or_02    1.138e+00  1.317e-01   8.645  < 2e-16 ***
Max_Number_Meds       2.515e-03  2.351e-04  10.699  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6947.8  on 5437  degrees of freedom
AIC: 6969.8

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7830  -1.0537  -0.1535   1.0924   1.7878

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.389e+00  3.022e-01  -4.595 4.32e-06 ***
Gender2               2.062e-01  6.181e-02   3.336 0.000851 ***
Language2             6.867e-01  9.729e-02   7.058 1.69e-12 ***
Marital_status2       3.725e-01  6.034e-02   6.173 6.70e-10 ***
Median_Income_by_Zip -4.994e-06  1.112e-06  -4.490 7.11e-06 ***
Lowest_Systolic_BP   -1.277e-02  1.964e-03  -6.502 7.95e-11 ***
Age                   2.711e-02  2.330e-03  11.636  < 2e-16 ***
time_since_diagnosis  1.287e-02  2.312e-03   5.567 2.59e-08 ***
Metoprolol_1_or_02   -2.598e-02  6.848e-02  -0.379 0.704361
bumetanide_1_or_02    1.244e+00  1.347e-01   9.238  < 2e-16 ***
Max_Number_Meds       2.379e-03  2.262e-04  10.521  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6862.3  on 5437  degrees of freedom
AIC: 6884.3

Number of Fisher Scoring iterations: 4
```

```
Call:
glm(formula = First_Admission ~ ., family = binomial(link = "logit"),
    data = training_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6896  -1.0603  -0.1491   1.1046   1.8298

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.448e+00  2.970e-01  -4.876 1.08e-06 ***
Gender2               3.262e-01  6.190e-02   5.271 1.36e-07 ***
Language2             7.118e-01  9.321e-02   7.636 2.24e-14 ***
Marital_status2       5.159e-01  6.011e-02   8.584  < 2e-16 ***
Median_Income_by_Zip -6.076e-06  1.134e-06  -5.357 8.46e-08 ***
Lowest_Systolic_BP   -1.263e-02  1.902e-03  -6.643 3.07e-11 ***
Age                   2.657e-02  2.356e-03  11.276  < 2e-16 ***
time_since_diagnosis  1.303e-02  2.416e-03   5.395 6.85e-08 ***
Metoprolol_1_or_02   -1.223e-01  7.000e-02  -1.748   0.0805 .
bumetanide_1_or_02    6.966e-01  1.369e-01   5.087 3.65e-07 ***
Max_Number_Meds       2.876e-03  2.267e-04  12.687  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7552.5  on 5447  degrees of freedom
Residual deviance: 6873.3  on 5437  degrees of freedom
AIC: 6895.3

Number of Fisher Scoring iterations: 4
```

Bibliography

[1] Massachusetts General Hospital, "Overview," 2017. [Online]. Available: http://www.massgeneral.org/about/overview.aspx. [Accessed 2 February 2017].

[2] Massachusetts General Hospital, "About Primary Care," 2017. [Online]. Available: http://www.massgeneral.org/primarycare/about/. [Accessed 1 April 2017].

[3] Massachusetts General Hospital, "Corrigan Minehan Heart Center Overview," 2017. [Online]. Available: http://www.massgeneral.org/heartcenter/about/overview.aspx. [Accessed 1 April 2017].

[4] U.S. News & World Report, "U.S. News Hospitals Rankings and Ratings," 2016. [Online]. Available: http://health.usnews.com/best-hospitals. [Accessed 5 February 2017].

[5] S. Dolcetti, "Analyzing the Impact of Delays for Patient Transfers from the ICU to General Care Units," Massachusetts Institute of Technology, Cambridge, 2015.

[6] W. Rieb, "Increasing Patient Throughput in the MGH Cancer Center Infusion Unit," Massachusetts Institute of Technology, Cambridge, 2015.

[7] A. Range, "Improving Surgical Patient Flow through Simulation of Scheduling Heuristics," Massachusetts Institute of Technology, Cambridge, 2013.

[8] V. Patel, "MGH Internal Medicine Associates – Primary Care Redesign," Massachusetts Institute of Technology, Cambridge, 2015.

[9] T. Schwartz, "Improving Surgical Patient Flow in a Congested Recovery Area," Massachusetts Institute of Technology, Cambridge, 2012.

[10] A. Marshall, "Improving Outpatient Non-Oncology Infusion through Centralization and Scheduling Heuristics," Massachusetts Institute of Technology, Cambridge, 2016.

[11] G. Lewis, "Overview of Heart Failure as a Complex Chronic Condition," Boston, 2016.

[12] S. Hunt, D. Baker and a. et, "ACC/AHA Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult: Executive Summary A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure)," *Circulation,* vol. 104, pp. 2996-3007, 2001.

[13] American Heart Association, "2016 Statistical Update," US Census Bureau, 2014 National Projections.

[14] J. Voigt, M. John et al., "A Reevaluation of the Costs of Heart Failure and its Implications for Allocation of Health Resources in the United States," *Clinical Cardiology,* vol. 37, no. 5, pp. 312-321, 2014.

[15] L. Allen and C. O'Connor, "Management of Acute Decompensated Heart Failure," *CMAJ,* vol. 176, no. 6, pp. 797-805, 2007.

[16] Centers for Medicare & Medicaid Services, "Readmissions Reduction Program (HRRP)," 18 April 2016. [Online]. Available: https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/readmissions-reduction-program.html. [Accessed 10 February 2017].

[17] Centers for Medicare & Medicaid Services, "Hospital Readmissions Reduction Program

(HRRP): Fiscal Year (FY) 2015 Fact Sheet," 1 July 2014. [Online]. Available: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HRRP/HRRP-2015-Fact-Sheet-.pdf. [Accessed 1 April 2017].

[18] M. R. Cowie, S. D. Anker et al., "Improving care for patients with acute heart failure: before, during and after hospitalization," *ESC Heart Failure,* vol. 1, no. 2, p. 110–145, December 2014.

[19] Johns Hopkins Medicine, "Basic Anatomy of the Heart," [Online]. Available: http://www.hopkinsmedicine.org/healthlibrary/conditions/cardiovascular_diseases/basic_anatomy_of_the_heart_85,P00192/. [Accessed 2 April 2017].

[20] K. Moore and E. Roth, "Heart Failure," Healthline Media, 9 December 2015. [Online]. Available: http://www.healthline.com/health/heart-failure#overview1. [Accessed 2 April 2017].

[21] American Heart Association, "Types of Heart Failure," 9 February 2017. [Online]. Available: http://www.heart.org/HEARTORG/Conditions/HeartFailure/AboutHeartFailure/Types-of-Heart-Failure_UCM_306323_Article.jsp#.WKuRoRiZNDU. [Accessed 10 February 2017].

[22] G. Couzens, "Heart Function, Not Failure," The New York Times, 29 January 2009. [Online]. Available: http://www.nytimes.com/ref/health/healthguide/esn-heart-failure-expert.html?pagewanted=all. [Accessed 11 February 2017].

[23] L. Polosajian, "Cardiovascular Services," MedNet Technologies, Inc., 2017. [Online]. Available: http://www.crsmdinc.com/services/cardiovascular-services/biventricular-pacemaker-implant/. [Accessed 2 April 2017].

[24] N. Albert, "A systematic review of transitional-care strategies to reduce rehospitalization in patients with heart failure," *Heart Lung,* vol. 45, no. 2, pp. 100-113, 2016.

[25] American College of Cardiology, "Hospital-to-Home (H2H). Improving transitions from the hospital to community settings.," 2013. [Online]. Available: https://cvquality.acc.org/~/media/QII/H2H/Getting%20Ready/1%20H2H%20Overview%20Slides.ashx. [Accessed 22 February 2017].

[26] American College of Cardiology, "Hospital-to-Home (H2H) Initiative," [Online]. Available: http://cvquality.acc.org/Initiatives/H2H.aspx. [Accessed 22 February 2017].

[27] M. Guazzi, K. Dickstein et al., "Six-minute walk test and cardiopulmonary exercise testing in patients with chronic heart failure: a comparative analysis on clinical and prognostic insights," *Circ Heart Fail,* vol. 2, no. 6, pp. 549-555, 2009.

[28] C. Soong, S. Daub et al., "Development of a checklist of safe discharge practices for hospital patients," *J Hosp Med,* vol. 8, no. 8, pp. 444-449, 2013.

[29] B. Wiggins, J. Rodgers et al., "Discharge counseling for patients with heart failure or myocardial infarction: a best practices model developed by members of the American College of Clinical Pharmacy's Cardiology Practice and Research Network based on the Hospital to Home (H2H) Initiative," *Pharmacotherapy,* vol. 33, p. 558–580, 2013.

[30] V. Paradis, S. Cossette et al., "The efficacy of a motivational nursing intervention based on the stages of change on self-care in heart failure patients," *J Cardiovasc Nurs,* vol. 25, pp. 130-141, 2010.

[31] M. White, R. Garbez et al., "Is “teach-back” associated with knowledge retention and hospital readmission in hospitalized heart failure patients?," *J Cardiovasc Nurs,* vol. 28, p. 137–146,

2013.

[32] T. Giardina, B. King et al., "Root cause analysis reports help identify common factors in delayed diagnosis and treatment of outpatients," *Health Aff (Millwood),* vol. 32, p. 1368–1375, 2013.

[33] M. Droppa and H. Lee, "Motivational interviewing: a journey to improve health," *Nursing,* vol. 44, pp. 40-45, 2014.

[34] M. Dart, Motivational Interviewing in Nursing Practice: Empowering the Patient, Sudbury, MA: Jones and Bartlett Publishers, LLC, 2011.

[35] The Joint Commission, "Advanced Certification Heart Failure Performance Measurement Implementation Guide," January 2014. [Online]. Available: http://www.jointcommission.org/assets/1/6/2014_ACHF_Manual_6_required_measures.pdf. [Accessed 20 February 2017].

[36] M. Gott, S. Barnes et al., "Patient views of social service provision for older people with advanced heart failure," *Health Soc Care Community,* vol. 15, p. 333–342, 2007.

[37] N. Albert, K. Trochelman et al., "Signs and symptoms of heart failure: are you asking the right questions?," *Am J Crit Care,* vol. 19, p. 443–452, 2010.

[38] P. Strachan, K. Currie et al., "Context matters in heart failure self-care: a qualitative systematic review," *J Card Fail,* vol. 20, p. 448–455, 2014.

[39] The Peak Performance Center, "Learning Pyramid," [Online]. Available: http://thepeakperformancecenter.com/educational-learning/learning/principles-of-learning/learning%20pyramid/. [Accessed 23 February 2017].

[40] S. Fredericks, H. Beanlands et al., "Effects of the characteristics of teaching on the outcomes of heart failure patient education interventions: a systematic review," *Eur J Cardiovasc Nurs,* vol. 9, pp. 30-37, 2010.

[41] L. Lemond and L. Allen, "Palliative care and hospice in advanced heart failure," *Prog Cardiovasc Dis,* vol. 54, pp. 168-178, 2011.

[42] S. Dev, A. Abernethy et al., "Preferences of people with advanced heart failure—a structured narrative literature review to inform decision making in the palliative care setting," *Am Heart J,* vol. 164, p. 313–319 e315, 2012.

[43] S. Agboola, K. Jethwani et al., "Heart Failure Remote Monitoring: Evidence From the Retrospective Evaluation of a Real-World Remote Monitoring Program," *Journal of Medical Internet Research,* vol. 17, no. 4, p. e101, 22 April 2015.

[44] A. S. Desai, "Home Monitoring Heart Failure Care Does Not Improve Patient Outcomes," *Circulation,* vol. 125, pp. 828-836, 13 February 2012.

[45] C. Murray, S. Agha et al., "The evaluation and monitoring of volume status in congestive heart failure.," *Congest Heart Fail.,* vol. 14, no. 3, pp. 135-140, May-Jun 2008.

[46] P. Adamson, W. Abraham et al., "CHAMPION trial rationale and design: the long-term safety and clinical efficacy of a wireless pulmonary artery pressure monitoring system.," *J Card Fail,* vol. 17, no. 1, pp. 3-10, 2011.

[47] A. Maisel, D. Barnard et al., "Primary Results of the HABIT Trial (Heart Failure Assessment With BNP in the Home)," *Journal of the American College of Cardiology,* vol. 61, no. 16, pp. 1726-1735, April 2013.

[48] J. Greenwald, P. Cronin et al., "A Novel Model for Predicting Rehospitalization Risk

Incorporating Physical Function, Cognitive Status, and Psychosocial Support Using Natural Language Processing," *Medical Care,* vol. 00, no. 00, pp. 1-6, 2016.

[49] P. Cronin, J. Greenwald et al., "Development and Implementation of a Real-Time 30- Day Readmission Predictive Model," Laboratory of Computer Science, Boston.

[50] W. C. Levy, D. Mozaffarian et al., "The Seattle Heart Failure Model: Prediction of Survival in Heart Failure," *Circulation,* vol. 113, pp. 1424-1433, 2006.

[51] P. Cronin, J. Greenwald et al., "Development and Implementation of a Real-Time 30- Day Readmission Predictive Model," Laboratory of Computer Science, Boston.

[52] U.S. Census Bureau, "Income Datasets," [Online]. Available: https://www.census.gov/topics/income-poverty/income.html. [Accessed 10 November 2016].

[53] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *Journal of Statistical Software,* vol. 45, no. 3, pp. 1-67, 2011.

[54] C. Boy, M. Tolson et al., "Evaluating trauma care: The TRISS method. Trauma Score and the Injury Severity Score," *The Journal of trauma,* vol. 27, no. 4, pp. 370-378, 1987.

[55] J. Truett, J. Cornfield et al., "A multivariate analysis of the risk of coronary heart disease in Framingham," *Journal of chronic diseases,* vol. 20, no. 7, pp. 511-524, 1967.

[56] K. Adams, C. Sueta et al., "Gender Differences in Survival in Advanced Heart Failure," *Circulation,* vol. 99, pp. 1816-1821, 1999.

[57] M. Wynia, C. Osborn et al., "Health Literacy and Communication Quality in Health Care Organizations," *J Health Commun.,* vol. 15, no. 2, p. 102–115, 2010.

[58] M. Chung, T. Lennie et al., "Marital status is an Independent Predictor of Event-Free Survival of Patients with Heart Failure," *Am J Crit Care,* vol. 18, no. 6, p. 562, 2009.

[59] D. G, S. E et al., "Influence of socioeconomic factors on hospital readmissions for heart failure and acute myocardial infarction in patients 65 years and older: evidence from a systematic review," *Clin Interv Aging,* vol. 10, pp. 237-45, 2015.

[60] M. Gheorghiade, W. Abraham et al., "Systolic Blood Pressure at Admission, Clinical Characteristics, and Outcomes in Patients Hospitalized With Acute Heart Failure," *JAMA,* vol. 296, pp. 2217-2226, 2006.

[61] University of Washington, "Seattle Heart Failure Model," 2006. [Online]. Available: https://depts.washington.edu/shfm/. [Accessed 17 March 2017].

[62] Partners HealthCare, "Hospitals and Affiliates," [Online]. Available: http://www.partners.org/Services/Hospitals-And-Affiliates.aspx. [Accessed 10 March 2017].

[63] Partners HealthCare, "Integrated Care Management Program," Partners HealthCare, [Online]. Available: http://www.partners.org/Innovation-And-Leadership/Population-Health-Management/Current-Activities/Integrated-Care-Management-Program.aspx. [Accessed 2 April 2017].

[64] A. Maisel, "B-Type Natriuretic Peptide Levels: Diagnostic and Prognostic in Congestive Heart Failure," *Circulation,* vol. 105, pp. 2328-2331, 21 May 2002.

[65] R Core Team, *R: A Language and Environment for Statistical Computing,* Vienna: R Foundation for Statistical Computing, 2013.