PREDICTING POPULATION HEALTH FOCUSED OUTCOMES USING MACHINE LEARNING

By

David Arnold, MSN, RN

A Dissertation Submitted to

Rutgers – School of Health Professions

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biomedical Informatics

Department of Health Informatics

School of Health Professions

Rutgers, the State University of New Jersey

February 2019

RUTGERS
School of Health Professions

Final Dissertation Defense Approval Form

Predicting Population Health Focused Outcomes Using Machine Learning

BY

David Arnold, MSN, RN

Dissertation Committee:

Shankar Srinivasan PhD

Frederick Coffman PhD

Suril Gohel PhD

Approved by the Dissertation Committee:

_____ **Date:**_____

_____ **Date:**_____

_____ **Date:**_____

_____ **Date:**_____

_____ **Date:**_____

TABLE OF CONTENTS

Abstract xvi

Chapter I 1

INTRODUCTION 1

Research Objectives: 1

Statement of Problem 2

Cost Distribution 3

Chronic Conditions and Cost 4

Background of the Problem 7

Significance of Research 8

Hypotheses 10

The random forest model will predict which patients will incur high cost greater than chance. 10

The random forest model will predict which patients will have at least one hospital admission greater than chance. 10

The random forest model will predict which patients will have at least one hospital readmission greater than chance. 10

The random forest model will predict which patients will have more than one emergency department (ED) encounter greater than chance. 10

Chapter II 11

LITERATURE REVIEW 11

Risk 11

Risk Stratification for Care Management 12
Data Used in Risk Adjustment 14
Demographics 14
Address 14
Diagnosis/ Condition 14
Clinical Features 15
Results 15
Vital Signs 15
Chronic Comorbid Indices 15
Charlson Comorbidity Index 16
Elixhauser Comorbidity Index 17
Risk Adjustment for Prospective Payment Adjustment 18
Diagnosis Related Groups 18
Resource Utilization Groups 18
Ambulatory Patient Groups 19
Medicare Severity Long-Term Care Diagnosis Related Groups 19
Home Health Resource Groups 19
Case Mix Groups 20
HHS Hierarchical Condition Categories 20
Claims-Based Risk Adjustment Models 20
3M Clinical Risk Groupers (CRG) 22
Milliman Advanced Risk Adjusters v3.6 26

Centers for Medicare and Medicaid Services HHS-HCC Model v3 27
Essential Differences Between Proposed Research and Prior Studies 29
Previous Work 29
Population Characteristics 30
Chapter III 32
METHODS 32
Data Acquisition 32
Patient Matching 32
Data 33
Conditions 33
Results 34
Procedure 34
Medication 34
Cost 34
Demographics 35
Claims 35
Electronic Medical Record 36
Hospital-based EMR 36
Ambulatory 37
Epic Ambulatory 37
GE Centricity EMR 37
Home Care 37

Long-Term Care and Rehab 38
Health Information Exchange 38
Jersey Health Connect 39
Patient Ping 39
Lab Data 39
LabCorp 40
Quest Diagnostics 40
Data Cleansing 40
Structure of the Data 41
Data Analysis: Random Forest 42
Classification Decision Trees 42
Gini Measure of Impurity 43
Many Features 43
Overfitting 44
Managing Noise in Data 44
Imbalanced Classes 44
Feature Selection and Curation 45
Feature Multicollinearity 45
Random Forest Model Performance 45
Receiver Operating Characteristic Curve (ROC) 47
Precision Recall Curve 49
Tuning Random Forest 49

Partitioning Data 50
Outcome Distributions 51
Chapter IV 52
RESULTS 52
Reporting Model Results 52
High Cost (U) 55
High Cost (C) 56
High Cost (P) 57
High Cost (M) 58
High Cost (R) 59
High Cost (U, C) 60
High Cost (U, C, P) 61
High Cost (U, C, P, M) 62
High Cost (U, C, P, M, R) 63
Admission (U) 64
Admission (C) 65
Admission (P) 66
Admission (M) 67
Admission (R) 68
Admission (U, C) 69
Admission (U, C, P) 70
Admission (U ,C, P, M) 71

Admission (U, C, P, M, R) 72
Readmission (U) 73
Readmission (C) 74
Readmission (P) 75
Readmission (M) 76
Readmission (R) 77
Readmission (U, C) 78
Readmission (U, C, P) 79
Readmission (U, C, P, M) 80
Readmission (U, C, P, M, R) 81
Multi-ED (U) 82
Multi-ED (C) 83
Multi-ED (P) 84
Multi-ED (M) 85
Multi-ED (R) 86
Multi-ED (U, C) 87
Multi-ED (U,C,P) 88
Multi-ED (U, C, P, M) 89
Multi-ED (U, C, P, M, R) 90
Chapter V 91
DISCUSSION 91
Limitations of the Data 92

Potential Sources of Error in Data 93

Considerations for Future Research 93

Chapter VI 95

SUMMARY AND CONCLUSION 95

LIST OF TABLES

Table 1 Charlson Comorbidity Conditions and Weights[36] 16

Table 2 Figure 7 Elixhauser Chronic Condition Categories 17

Table 3 Clinical Risk Grouper Categories 24

Table 4 CRG Core Health Status (Table 3)[56] 25

Table 5 MARA and Charlson Comorbidity Index Blended Model[7] 30

Table 6 Study Population Race Distribution 31

Table 7 Cost Distribution 35

Table 8 Hospitals Contributing Data to EDW 36

Table 9 Data Structure for Random Forest 41

Table 10 Error Matrix 46

Table 11 Outcome Distributions Full Dataset 51

Table 12 High Cost (U) Test Error Matrix 55

Table 13 High Cost (C) Test Error Matrix 56

Table 14 High Cost (P) Test Error Matrix 57

Table 15 High Cost (M) Test Error Matrix 58

Table 16 High Cost (R) Test Error Matrix 59

Table 17 High Cost (U,C) Test Error Matrix 60

Table 18 High Cost (U,C,P) Test Error Matrix 61

Table 19 High Cost (U,C,P,M) Test Error Matrix 62

Table 20 High Cost (U, C, P, M, R) Test Error Matrix 63

Table 21 Admission (U) Test Error Matrix 64

Table 22 Admission (C) Test Error Matrix....65
Table 23 Admission (P) Test Error Matrix....66
Table 24 Admission (M) Test Error Matrix....67
Table 25 Admission (R) Test Error Matrix....68
Table 26 Admission (U, C) Test Error Matrix....69
Table 27 Admission (U, C, P) Test Error Matrix....70
Table 28 Admission (U ,C, P, M) Test Error Matrix....71
Table 29 Admission (U, C, P, M, R) Test Error Matrix....72
Table 30 Readmission (U) Test Error Matrix....73
Table 31 Readmission (C) Test Error Matrix....74
Table 32 Readmission (P) Test Error Matrix....75
Table 33 Readmission (M) Test Error Matrix....76
Table 34 Readmission (R) Test Error Matrix....77
Table 35 Readmission (U,C) Test Error Matrix....78
Table 36 Readmission (U, C, P) Test Error Matrix....79
Table 37 Readmission (U, C, P, M) Test Error Matrix....80
Table 38 Readmission (U, C, P, M, R) Error Matrix....81
Table 39 Multi-ED Test Error Matrix....82
Table 40 Multi-ED Test Error Matrix....83
Table 41 Multi-ED (P) Test Error Matrix....84
Table 42 Multi-ED Test Error Matrix....85
Table 43 Multi-ED (R) Test Error Matrix....86

Table 44 Multi-ED (U,C) Test Error Matrix.....87

Table 45 Multi-ED (U,C,P) Error Matrix 88

Table 46 Multi-ED (U, C, P, M) Test Error Matrix 89

Table 47 Multi-ED (U, C, P, M, R) Test Error Matrix.....90

LIST OF FIGURES

Figure 1 Distribution of Healthcare Utilizers in Population[6 7] 4

Figure 2 Distribution of Cost by Number of Chromic Conditions[7,10] 5

Figure 3 Cost Distribution by Number of Chronic Conditions[7,10] 6

Figure 4 ED Visits per 1,000 Beneficiaries by Number of Chronic Conditions[7,8] 6

Figure 5 Hospital Readmissions Percentage by Number of Chronic Conditions[7,8] 7

Figure 6 Study Population Age Distribution....31

Figure 7 Visualizing the Gini Index 43

Figure 8 ROC Plot 48

Figure 9 Error Rates with Two Thousand Trees 50

Figure 10 Model Comparison 54

Figure 11 High Cost (U) Test Curves....55

Figure 12 High Cost (C) Test Curves 56

Figure 13 High Cost (P) Test Curves 57

Figure 14 High Cost (M) Test Curves 58

Figure 15 High Cost (R) Test Curves 59

Figure 16 High Cost (U,C) Test Curves 60

Figure 17 High Cost (U,C,P) Test Curves 61

Figure 18 High Cost (U,C,P,M) Test Curves....62

Figure 19 High Cost (U, C, P, M, R) Test Curves 63

Figure 20 Admission (U) Test Curves....64

Figure 21 Admission (C) Test Curves 65

Figure 22 Admission (P) Test Curves 66
Figure 23 Admission (M) Test Curves 67
Figure 24 Admission (R) Test Curves 68
Figure 25 Admission (U, C) Test ROC Curve 69
Figure 26 Admission (U,C,P) Test Curves 70
Figure 27 Admission (U ,C, P, M) Test Curves 71
Figure 28 Admission (U, C, P, M, R) Test Curves 72
Figure 29 Readmission (U) Test Curves 73
Figure 30 Readmission (C) Test Curves 74
Figure 31 Readmission (P) Test ROC Curve 75
Figure 32 Readmission (M) Test Curves 76
Figure 33 Readmission (R) Test ROC Curve 77
Figure 34 Readmission (U,C) Test Curves 78
Figure 35 Readmission (U, C, P) Test Curves 79
Figure 36 Readmission (U, C, P, M) Test ROC Curve 80
Figure 37 Readmission (U, C, P, M, R) Test ROC Curve 81
Figure 38 Multi-ED (U) Test Curves 82
Figure 39 Multi-ED (C) Test Curves 83
Figure 40 Multi-ED Test Curves 84
Figure 41 Multi-ED (M) Test Curves 85
Figure 42 Multi-ED (R) Test ROC Curve 86
Figure 43 Multi-ED (U,C) Test ROC Curve 87

Figure 44 Multi-ED (U,C,P) Test Curves 88

Figure 45 Multi-ED (U, C, P, M) Test Curves 89

Figure 46 Multi-ED (U, C, P, M, R) Test ROC Curve 90

Abstract

Care management activities seek to reduce healthcare cost and improve patient outcomes. Identifying patients who may receive substantial benefit from care management services can be especially challenging when managing large populations across disparate systems. This research tests a novel method for identifying patients for care management using over 30 disparate healthcare data sources and machine learning. Random Forest models were used to predict four binary outcomes; high cost, hospital admission, hospital readmission, and multiple emergency department visits. The models leveraged population health enterprise data warehouse cross-ontology mappings for the following data types; conditions, procedures, medications, results, demographics, and claims-based cost and utilization. Each of the data types were tested independently then combined incrementally. The highest performing models for each outcome of interest resulted with the following ROC AUC; High Cost (0.81), Admission (0.80), Re-admission (0.86), and Multi-ED (0.74). The research shows disparate data sources and machine learning can be used to predict population health focused outcomes. The framework used in this research has the potential to expand and scale to include any number of additional data types and outcomes.

Chapter I

INTRODUCTION

The purpose of this research is to test a novel method for identifying patients who may benefit from care management using disparate healthcare data sources and machine learning. The research uses a unique and novel approach leveraging a dataset from an enterprise data warehouse developed to support an integrated delivery network. The database consists of over 30 disparate data sources that have been normalized and standardized into a person-centered platform with over 4.4 million unique patients and over 14,000 mapped clinical concepts. The dataset along with machine learning were leveraged to predict four binary outcomes of interest for care management activities.

Research Objectives:

a. Predict which patients will incur high cost in a future calendar year.
b. Predict which patients will have at least one hospital admission in a future calendar year.

c. Predict which patients will have at least one hospital readmission in a future calendar year.
d. Predict which patients will have more than one emergency department (ED) encounter in a future calendar year.

Statement of Problem

Healthcare cost in the United States is approaching 20% of the nation's gross domestic product while the quality of care is below many other developed countries resulting in low healthcare value[1]. In an effort to improve value, reimbursement mechanisms for healthcare services are transforming from the fee-for-service model to a variety of value-based models. Health systems are beginning to manage the total cost and quality for patient populations under these new value-based models. These changes are shifting the financial responsibilities to the providers in whole or in part and removes the financial incentive to increase utilization[2]. To simply reduce treatment lowering utilization and cost would be inappropriate. Utilization and cost should be reduced while simultaneously improving quality or outcomes. This balance must be monitored and maintained by those implementing reimbursement transformation. It has been estimated that $700 billion is wasted annually in unnecessary utilization of healthcare in the US[3]. Coordination of care and care management are areas that seek to reduce waste and improve quality for patient populations. Targeting patients that could benefit most from care management and care coordination services is challenging due to the volume of patients and the complex nature of healthcare data. There is little research on the risk

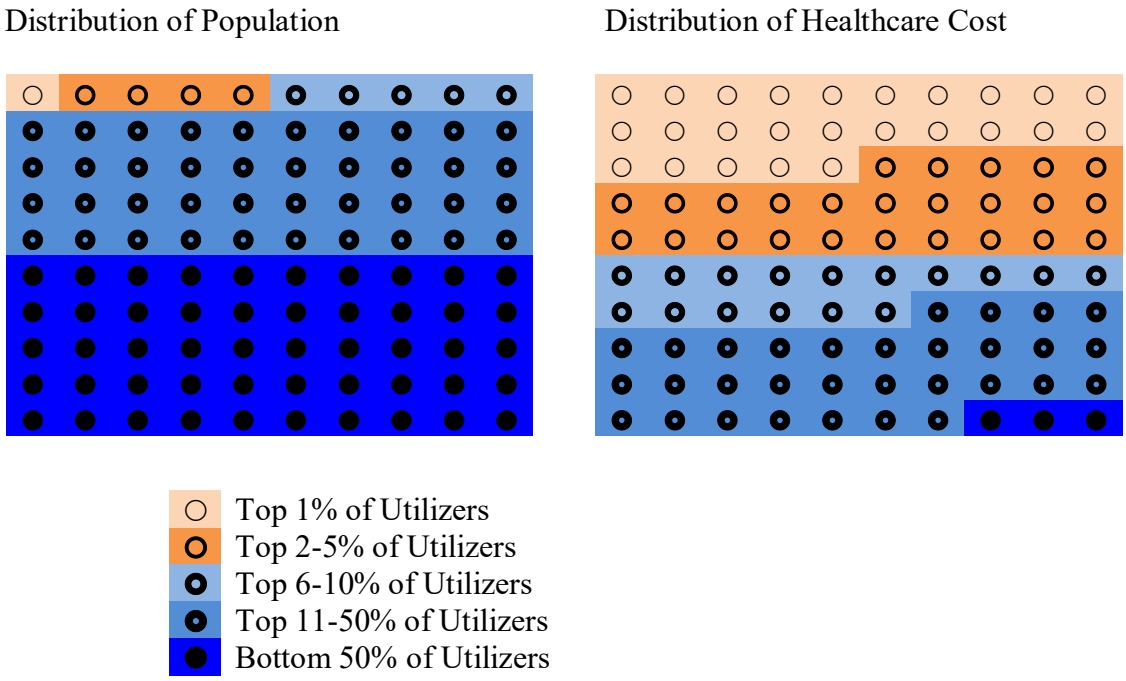
stratification of populations for the purposes of care management. This dissertation will use the Hackensack Meridian Health population health enterprise data warehouse and machine learning to identify patients who are likely to incur high future cost, inpatient admission, inpatient readmission, and multiple ED encounters.

Hackensack Meridian Health (HMH) is an integrated delivery network in New Jersey comprised of a diverse portfolio of healthcare services including 13 acute-care hospitals, a clinically integrated network, over 3,000 employed physicians, over 7,000 affiliated physicians, home care, long-term care and rehab, ambulance, urgent care, retail care, surgery centers, and more. This research will focus on the risk stratification of the HMH Clinically Integrated Network (CIN) patient population for the purposes of care management. The CIN is made up of employed and independent physicians who are contracted with commercial insurers as an Accountable Care Organization (ACO). The following basic theories pertain to the proposed research:

Cost Distribution

Healthcare cost is not normally distributed across populations. Vilfredo Pareto's economic theory now called *the Pareto Law*, noted 80 percent of the wealth is held by 20 percent of the population[4,5]. Similar to wealth, healthcare cost is also not distributed evenly in the U.S. The top one percent of patients are responsible for roughly 25 percent of expenditures, the top five percent account for roughly 50 percent of expenditures, the top ten percent are responsible for 66 percent, and the top half are responsible for 97 percent of expenditures[6]. Figure 2 shows a depiction of the portions of U.S. population and their proportion of healthcare spending

Figure 1 Distribution of Healthcare Utilizers in Population[6] [7]



Chronic Conditions and Cost

The prevalence of chronic disease has increased in recent years and the percentage of Americans living with chronic disease is projected to reach 50% by 2030[8]. Advances in medical treatment are allowing patient's with chronic disease to live longer lives[8]. Eighty three percent of patients with Medicare have at least one chronic condition and those with at least five chronic conditions see an average of 13 physicians and fill 50 prescriptions annually[9]. These complex care scenarios are ripe for waste and breakdown of care coordination.

Chronic conditions are correlated to cost. Generally, as the number of chronic conditions increases the cost increases [10]. Additionally patient healthcare utilization

patterns increase along with the number of chronic conditions[10]. Multiple chronic conditions are a strong predictor of high future cost[10,11].

Figure 2 Distribution of Cost by Number of Chromic Conditions[7,10]

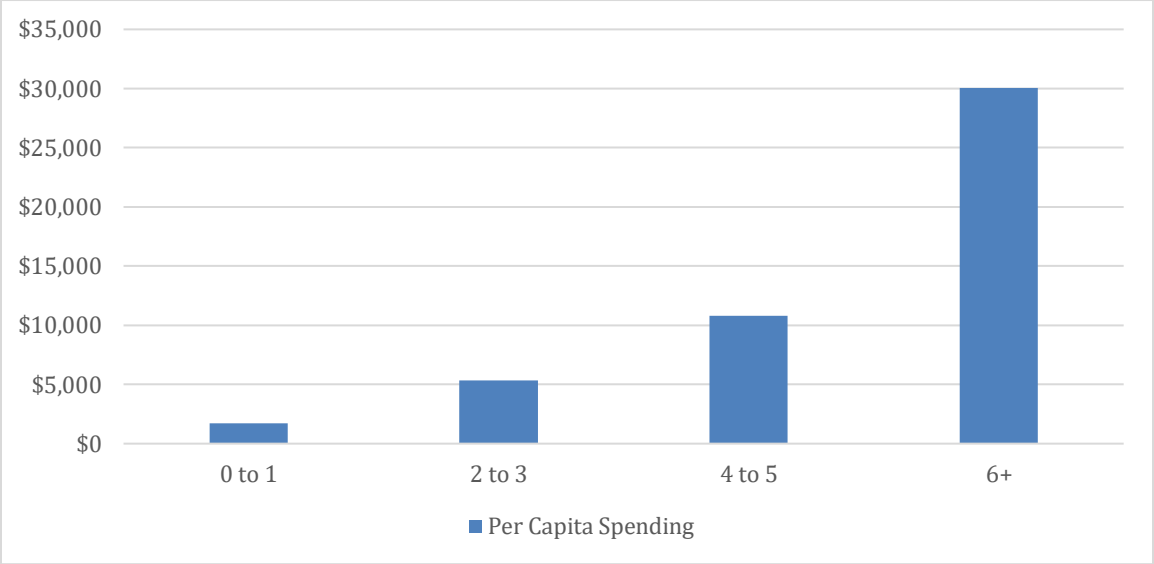


Figure 3 Cost Distribution by Number of Chronic Conditions[7,10]

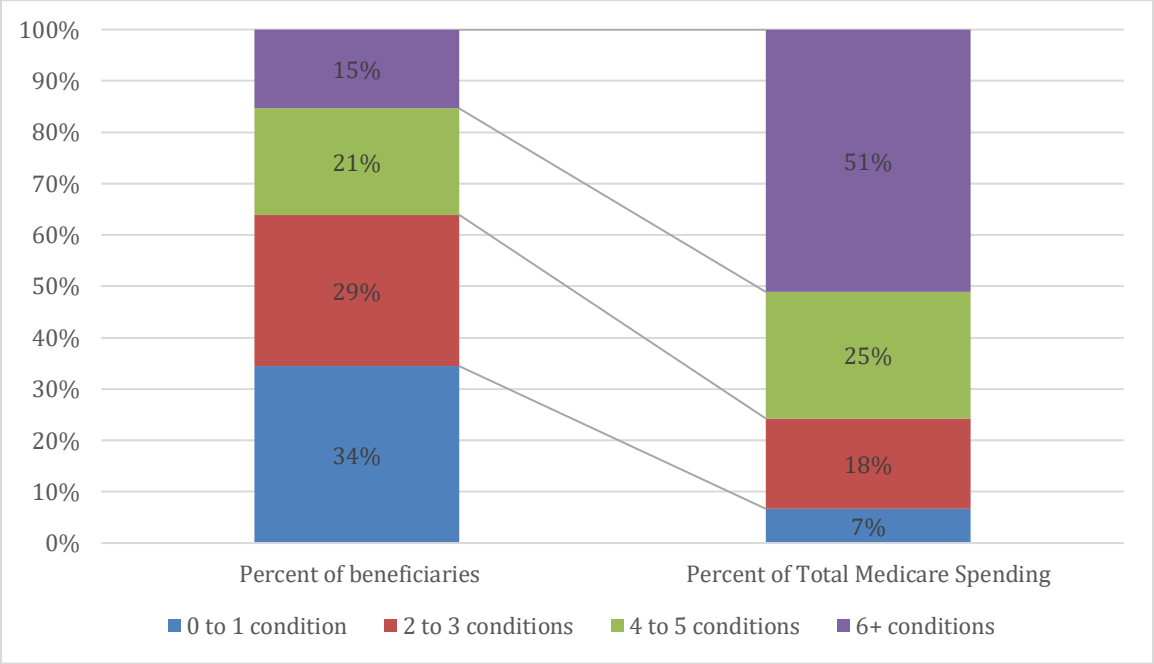


Figure 4 ED Visits per 1,000 Beneficiaries by Number of Chronic Conditions[7,8]

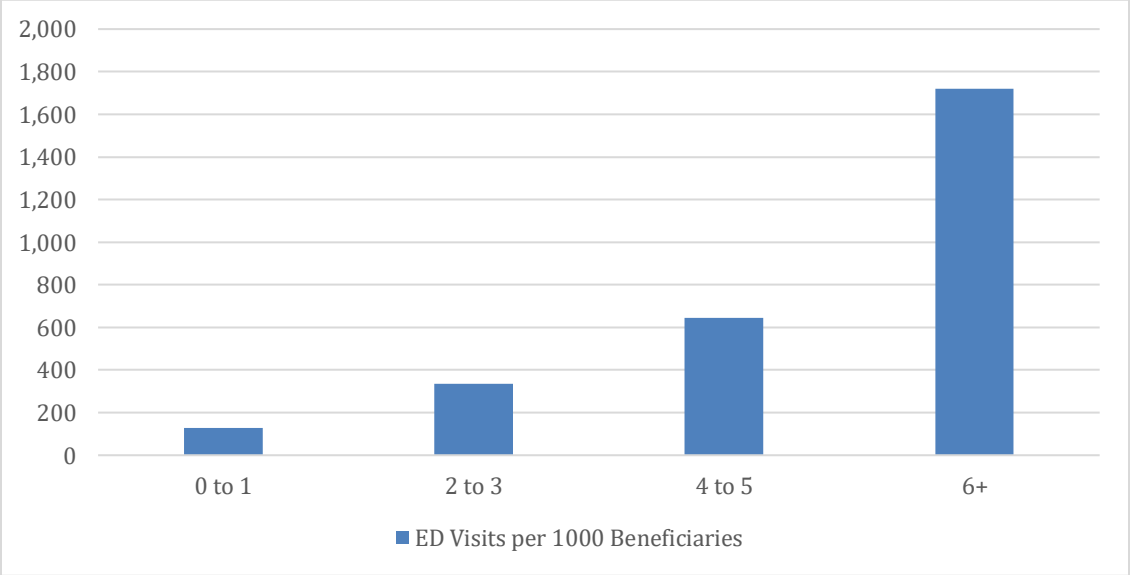
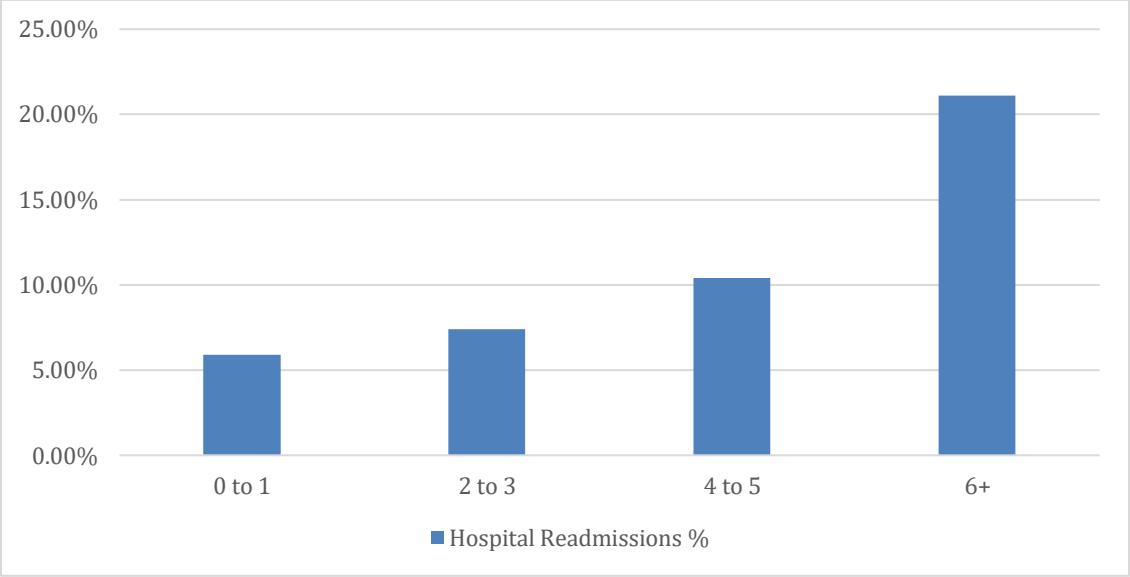


Figure 5 Hospital Readmissions Percentage by Number of Chronic Conditions[7,8]



Background of the Problem

While under the fee-for-service reimbursement model, the cost of healthcare in the United States has become among the highest in the world while measures of value are consistently below those of other similarly developed countries[1]. To address this problem, the Institute for Healthcare Improvement proposed the now widely accepted goal of the *triple aim* of improving patient experience, improving clinical outcomes, and lowering cost per-capita[12]. Achieving the triple aim in the United States is a national priority with broad bi-partisan political support[13].

As new reimbursement models are implemented, health systems and healthcare providers are being compelled to change their delivery frameworks and business practices to increase value. Predictive and preventive care approaches are one of the key areas of focus. These approaches can be successful through the use of information technology and analytics. For many years the field of Actuarial Science has developed claims-based models to predict future medical cost. While the tools are valuable, they are only able to predict cost accurately 20-40% of the time and most are costly and are proprietary in nature [14]. While the models have improved over time, they require adjudicated claims and most must be licensed from vendors. Adjudicated claims are not always available for populations being treated. Risk stratification models that can function with or without claims would be beneficial for treating the entire population being served. This is a common challenge in the current state of healthcare transformation.

Significance of Research

This research seeks to improve an important social and economic problem of high total cost of healthcare. This research can benefit healthcare as health systems continue to consolidate and move towards value-based reimbursement models, effective identification of patients who could benefit from care management services is important[15]. This research can also demonstrate a successful method for leveraging data assets from an integrated delivery network to target patients for intervention that can be used by other organizations. This research is important because the use of both clinical

and administrative data to risk stratify patients has been posed by others and promise has been shown in the use of machine learning for risk stratifying patients[14,16].

This research is significant to the field of biomedical informatics because it is driven by diverse voluminous healthcare data. It can demonstrate the use of big data and machine learning for identifying patients for care management. The research relies heavily on healthcare ontologies and concept mappings between those ontologies both of which are core to biomedical informatics. The integration of disparate healthcare data sources is a challenge for many researchers and healthcare systems. Most population risk stratification methods have relied on demographics and diagnosis data, this research will expand to other clinical data. Real world use of both big data platforms and machine learning are areas of interest for both the healthcare industry and the science of biomedical informatics.

Hypotheses

The random forest model will predict which patients will incur high cost greater than chance.

The random forest model will predict which patients will have at least one hospital admission greater than chance.

The random forest model will predict which patients will have at least one hospital readmission greater than chance.

The random forest model will predict which patients will have more than one emergency department (ED) encounter greater than chance.

Chapter II

LITERATURE REVIEW

This literature review will focus on the healthcare domains of risk measurement and care management and their synthesis. The major types of risk measurement and adjustment tools are reviewed to demonstrate prior work in this area and to lay a framework for this research.

Risk

The term *risk* has a number of interrelated meanings. In the context of healthcare, it can refer to the possibility of loss or injury, or an unwanted health outcome. In the context of health insurance it can refer more specifically to financial loss[17]. In this research, the term risk refers to both of these definitions. Iezzoni noted that when calculating any type of risk, it is important to answer the following questions:

- What is the purpose of identifying risk?
- What is the population?
- What is the outcome for which risk is being calculated? In other words, "Risk of what?"

- What is the timeframe being used to calculate risk? [18]

Risk adjustment is an activity used to compare outcomes across populations, providers, treatments, geographies, etc., while accounting for variability in patient-based factors[19]. Patients with advanced disease are often prone to medical complications and sequelae resulting in unwanted outcomes. For this reason outcomes measurement is often risk adjusted in an effort to make comparisons more equitable[19]. There are a number of factors that influence various patient outcomes. "Outcomes = *f* (intrinsic patient-related risk factors, treatment effectiveness, quality of care, random chance)."[19] In the context of care management, quality of care and the coordination of treatment are areas that can influence outcomes. Risk adjustment and risk prediction are commonly attempted in tandem as the results can be used for both purposes[14]. Due to this close relationship and the goal of this research, Risk Adjustment methods are included in the following literature review.

One key area that employs risk prediction is the health insurance industry through actuarial science. Accurately predicting future cost is a key for the successful operation of an insurance plan. When historical claims are available, insurers use this data along with demographic information to predict future cost and set premium payments. Prospective payment models are another area where risk adjustment has been used.

Risk Stratification for Care Management

Identifying patients who are projected to incur higher costs in an effort to improve outcomes and efficiency are typical goals of care management activities[15,16,20-22].

Care management relies on risk stratification for the following reasons:

- Resources used to deliver care management are typically limited
- Not all patients benefit similarly from care management
- Patient needs for care management may change over time
- Care management is needed across all venues of care

This can be summed up by stating "*for who, when, and where do care management teams focus activities?*"

The methods used in identifying patients for care management programs can be grouped into three categories, quantitative, qualitative, and hybrid[20]. Quantitative models are typically regression-based and use data to predict outcome of interest related to care management. Most commercially available population-based risk adjustment tools are quantitative and repurposed actuarial instruments used to predict total expenditure [14]. Other tools predict outcomes like resource use, mortality, hospital admissions, readmissions, length of inpatient hospital stay, disease complexity, and organ system involvement[18].

Qualitative methods include patient self-referral and clinician referral as well as qualitative surveying[20]. Hybrid models are a combination of qualitative and quantitative identification of high-risk patients for care management. This research falls into the quantitative method category.

Data Used in Risk Adjustment

Administrative data is commonly used in healthcare research as well as risk analysis. Administrative data includes healthcare billing data, adjudicated claims, demographics, and other types of data that are routinely collected during the regular operation of healthcare business. These data sets are often large, representative of care across settings and over time, are frequently de-identified [19].

Demographics

Patient demographics are commonly used in risk adjustment. Variables include age, sex, race and ethnicity [23]. In addition to demographics, marital status, primary language and country of origin have been used in risk adjustment[19]. Age is associated with increased number of chronic conditions and cost.

Address

Patient zip code has been correlated to patient outcomes [24-27]. The Public Health Disparities Geocoding Project found census tract to be a more suitable data type for health disparities however census tract data was not available for the patient population in this study [28]. The first three digits of the patients' latest zip code from payer sources will be used in this research.

Diagnosis/ Condition

Diagnosis and medical condition or problem data is used in most risk adjustment or actuarial models. Any active diagnosis recorded in a contributing data source during the baseline measurement year will be included in the data set.

Clinical Features

Clinical data adds significant explanatory power to cost-focused risk models [29]. The clinical data can be categorized as follows

Results

Results in this context refer to laboratory results. The results are mapped from the various contributing sources. Lab results are not typically available in administrative data sets. The claim or bill for the lab would be present but not the diagnostic result. One challenge with results are the various reference ranges used across diagnostic machines. This challenge will be evaluated during data analysis. The results from the latest date during the measurement year will be used when more than one result value is present for a given result.

Vital Signs

Vital signs are included in the result extracts from EMR and HIE sources. The vital signs from the latest date within the baseline 2016 calendar year will be used for this research.

Chronic Comorbid Indices

As discussed earlier in this paper, chronic conditions are correlated to higher cost and higher utilization. A number of well-established comorbidity indices have been used to risk adjust and predict future cost[30-33].

Charlson Comorbidity Index

The Charlson Co-morbidity Index was developed to predict one-year mortality in cancer patients[34]. The index identifies 17 chronic conditions and applies a cumulative weighted score based on those conditions resulting in a index score for each patient[34]. It was used subsequently repurposed to identify patients who will incur high future cost and produced an R^2 value of 0.25 [31,35]. The predictive power of the Charlson Index has been demonstrated in both general and disease specific populations[30]. It has performed similarly compared to other clinically –based risk algorithms due to the correlation between mortality and the other dependent variables of interest in other studies[22,31]. The original published methods relied on chart abstraction but others have adapted the index for use with administrative, claims data. The conversion from clinical abstraction-based diagnosis to ICD-9 diagnosis codes was first published by Deyo, Cherkin, and Ciol[36]. Later it was translated to ICD-10 by Hude Quan et al.[36,37].

Table 1 Charlson Comorbidity Conditions and Weights[36]

Chronic Disease	Weight	Chronic Disease	Weight	Chronic Disease	Weight
Tumor	2	Pulmonary	1	Hiv/Aids	6
Heart Failure	1	Mild Liver Disease	1	Severe Liver Disease	3
Myocardial Infarction	1	DM without Complications	1	Peptic	1
Peripheral Vascular Disease	1	DM with Complications	1	Hemiplegia	2
Cerebrovascular Disease	1	Metastatic Tumor	6	Connective Tissue	1
				Renal	2

Elixhauser Comorbidity Index

The Elixhauser Comorbidity Index was developed to risk adjust for medical and financial outcomes using administrative diagnosis data[38]. Researchers began with a list of 41 comorbidities gleaned from literature[38]. These were analyzed for frequency and statistical significance and some heterogeneous comorbidities were further subdivided resulting in a model with 30 chronic condition categories[38]. Nearly all of the chronic condition categories were found to be positively correlated to high cost[38].

Table 2 Figure 7 Elixhauser Chronic Condition Categories

1. Congestive Heart Failure
2. Cardiac Arrhythmias
3. Valvular Disease
4. Pulmonary Circulation Disorders
5. Peripheral Vascular Disorders
6. Hypertension
7. Paralysis
8. Other Neurological Disorders
9. Chronic Pulmonary disease
10. Diabetes, Uncomplicated
11. Diabetes, Complicated
12. Hypothyroidism
13. Renal Failure
14. Liver Disease
15. Peptic Ulcer Disease Excluding Bleeding
16. AIDS
17. Lymphoma
18. Metastatic Cancer
19. Solid Tumor Without Metastasis
20. Rheumatoid Arthritis
21. Coagulopathy
22. Obesity
23. Weight Loss
24. Fluid and Electrolyte Disorder
25. Blood Loss Anemia
26. Deficiency Anemias
27. Alcohol Abuse
28. Drug Abuse
29. Psychoses
30. Depression[38]

Risk Adjustment for Prospective Payment Adjustment

Diagnosis Related Groups

The US Department of Health and Human Services uses a number of risk adjustment tools to adjust prospective payments made to healthcare providers and insurers. The Diagnosis-Related Groups (DRG) use a method called "*the significant attribute method"* which divides diagnosis codes into broad disease areas and then into sub-categories grouped by typical or expected length of stay[39]. Initially there were 83 major diagnostic categories and 383 DRGs[39]. The DRG system was first used in 1980 in New Jersey to prospectively pay for hospital care using a weighted scale with 470 unique DRGs[40]. The DRG system underwent enhancements over the years with accompanying re-naming to Refined DRGs (RDRGs), All-Patient DRGs (APDRGs), Severity DRGs (SDRGs), All-Patient Refined DRG (APR-DRGs), and Medicare Severity DRGs (MS-DRGs)[41,42].

Resource Utilization Groups

Resource Utilization Groups (RUGs) were created to group nursing home residents into similar utilization groups[43,44]. Initially 44 categories were used with data collected through the Minimum Data Set (MDS), the Triggers and Resident Assessment Protocol (RAPs), and Utilization Guidelines[45]. The 66 RUGs included in version four are used to adjust prospective payments for Skilled Nursing Facilities (SNF) based on the number of minutes of physical, occupational, and speech therapy required for SNF residents[44].

Ambulatory Patient Groups

In 1986 the US Congress passed the Omnibus Budget Reconciliation Act Ambulatory Payment Classification (APC) charging the US Department of Health and Human Services to develop a prospective payment system for hospital outpatient services which had grown rapidly since the implementation of the inpatient DRG prospective payment system in the early 80's[46]. The Ambulatory Patient Groups (APGs) were developed to account for the type and volume of resources used in ambulatory encounters[46]. The APGs are identified by splitting treatment into three mutually exclusive groups; Significant Procedure, Medical Treatment, and Ancillary Services[46]. Instead of using diagnosis as the first classifier, APGs use procedures. The initial version of the APGs had 289 groups, version 3.0 had 478 groups[47].

Medicare Severity Long-Term Care Diagnosis Related Groups

The Medicare Severity Long-Term Care DRG (MS-LTC-DRGs) use the MS-DRGs with weighting calibrated for utilization in Long-Term Care (LTC) facilities [48]. The MS-LTC-DRGs are used to provide Medicare prospective payments in LTC setting.

Home Health Resource Groups

Similarly for home health services, a prospective payment model is used for Medicare patients. In the home care setting the Outcome and Assessment Information Set (OASIS) is completed and used to classify patients into one of 153 Home Health Resource Groups (HHRG) defined by hierarchical ranking in three categories; clinical, functional, and service[49].

Case Mix Groups

Inpatient Rehabilitation Facilities (IRF) use Case Mix Groups (CMGs) for Medicare prospective payments. These are calculated first by identifying the Rehabilitation Impairment Category based on the primary condition driving the admission. Patients are then grouped by their age, functional motor and cognitive status. Within each CMG patients are categorized further into four tiers based on co-morbidities[50].

HHS Hierarchical Condition Categories

The Health and Human Services Hierarchical Condition Categories (HCCs) are used to risk adjust Medicare capitation payments for beneficiaries enrolled in Medicare Part-C. In 2017 there were over 19.5 million beneficiaries enrolled in a managed Medicare plan, a 10-year growth rate of 95% in Medicare Part-C enrollment rate compared to total Medicare beneficiaries[51]. The HCC Model uses demographics and diagnosis data to assign patients a risk score. The model was selected by Medicare to adjust payments for managed Medicare (Medicare Part C) patients in an effort to combat patient selection bias [29]. The HCC model will be further reviewed in the risk adjustment models section of this paper.

Claims-Based Risk Adjustment Models

Claims-based quantitative risk adjusting models have become more prevalent in recent years. These models use demographic information, diagnosis, pharmacy utilization, and prior cost information to predict future total medical expenditure for

patient populations. The performance of the models has improved over time. Early models included only age and gender and produced an R^2 value less than 2%[52]. It is important to note that these models do not use clinical data from medical records[14]. The claims-based models typically follow a general process of categorizing medical codes into homogenous groups then calculate relative risk scores based on these classifications for individual patients and populations[53]. Commercially available models typically include concurrent versions that are explanatory in nature and are often used to risk adjust performance or for benchmarking. These models use the data to explain cost variation within the measurement period. Prospective models look at historical data to predict future cost. It is important to use these models appropriately. For purposes of predicting which patients would benefit most from care management the prospective type of model would be most appropriate.

The most recent assessment of claims-based risk adjustment models published by the Society of Actuaries included 23 prospective models and 19 concurrent models from 11 vendors and these models had R^2 values ranging between 15 and 28[14]. One of the advantages of claims-based risk models are the ability to capture all of the billed encounters, procedures, tests, and prescription medications across all venues of care for a given population[20]. However, clinical information such as labs and assessment findings are not present in this data. For example, two patients may be diagnosed with diabetes, one with a hemoglobin A1C level of 5.2, and the other with a hemoglobin A1C level of 11.6. The clinical risk associated with these lab values may be very different however

both patients may be coded with a diagnosis of diabetes and have a claim for a lab test. In this example the result value provides greater insight to the patients clinical state.

While claims data provides a broad view of healthcare provided for a given patient and population, it does not provide the deep clinical information found in clinical systems such as an electronic medical record. These systems typically contain rich clinical information but the information is typically less broad as compared to claims, unless a patient has received all of their medical care within one specific EMR system. Claims data has also been criticized for missing self-pay or uninsured patient data and for inherent weakness in providing a means for tracking patients who move between private insurers or between private and public coverage[19].

For the purposes of this review, only the following models in use at HMH will be included. As previously mentioned, there are more than 23 models currently available in the market[14].

3M Clinical Risk Groupers (CRG)

The development of the Clinical Risk Groupers (CRG) model was initially funded by the Department of Commerce's National Institutes of Standards and Technology, Advanced Technology Program, with three objectives in mind; a) to develop a clinically meaningful tool for predicting health care expenditures, b) to develop a tool for risk adjusting capitated payments, c) to develop a bridge between clinical and financial elements of health care [54]. The model uses diagnosis, procedure, demographics, pharmacy, and functional health status (when available) to assign patients to one of 1,408 mutually exclusive CRG's[54,55]. The CRG model includes four to six Severity of Illness

(SOI) levels within the categories[56]. This provides a greater level of granularity as to the burden of the disease and potential cost.

Step one creates a profile by identifying Major Diagnostic Categories (MDC) and Episode Diagnostic Categories (EDC) for the patient population. Step two identifies the Primary Chronic Disease (PCD) within the MDC and identifies the severity levels within each PCD. A single diagnosis can only be used as a severity modifier in one PCD it cannot be used to adjust severity in any other MDC and the severity adjustment is made to the highest ranking EDC.[56]

Step three assigns each patient to one of nine Core Health Status Ranks representing a scale from healthy to catastrophic.[56]

Table 3 Clinical Risk Grouper Categories

Categories	Count
1. Major Diagnostic Category (MDC)	37
a. Episode Diagnostic Category (EDC)	534
i. Chronic	164
(Lifelong or prolonged)	
1. Dominant	59
a. 4 levels of severity	
2. Moderate	65
a. 4 levels of severity	
3. Minor Chronic	40
a. 2 levels of severity	
ii. Acute	264
(Short, self-resolving or curative treatment exists)	
1. Significant Acute	156
2. Minor Acute	108
iii. Manifestations of Chronic Disease	106

Table 4 CRG Core Health Status (Table 3)[56]

Core Health Status	Examples
1. Healthy No chronic diseases and no significant acute illness in the past 6 months	
2. History of significant acute disease No PCD but at least 1 significant acute disease occurred in most recent 6 months	Pneumonia, pancreatitis, pelvic inflammatory disease
3. Single minor chronic disease Only 1 minor PCD	Migraine, chronic stomach ulcer
4. Minor chronic disease in multiple systems 2 or more minor PCDs	Chronic bronchitis and benign prostatic hypertrophy, migraine and hyperlipidemia
5. Single dominant or moderate chronic disease Only 1 dominant or moderate chronic PCD	CHF, diabetes, cerebrovascular disease, asthma
6. Significant chronic disease in multiple organ systems Identified by the presence of 2 or more PCDs of which at least 1 is a dominant or moderate chronic disease (but no more than 2 dominant chronic PCDs; minor PCDs that are at severity level 2 or higher are considered significant chronic diseases, but PCDs that are severity level 1 minor chronic disease are not used in this status level	CHF and cerebrovascular disease Diabetes and 1 other dominant chronic disease
7. Dominant chronic disease in three or more organ systems Dominant chronic PCDs in 3 or more organ systems	CHF and diabetes and COPD CHF and 2 or more other dominant chronic diseases
8. Dominant and metastatic malignancies Include primary malignancies that dominate the medical care required, or a nondominant malignancy that is metastatic (nondominant or nonmetastatic malignancies are treated as moderate chronic diseases)	Lung cancer, stomach cancer, metastatic prostate cancer
9. Catastrophic conditions Includes long-term dependency on a medical technology (eg, dialysis, respirator, total parenteral nutrition) and life-defining chronic diseases or conditions that dominate the medical care required	Dependence on dialysis, ventilator dependence, persistent vegetative state

Milliman Advanced Risk Adjusters v3.6

The Milliman Advanced Risk Adjusters (MARA) uses demographic data, medical, and pharmaceutical claims to predict healthcare utilization. Both concurrent and prospective models are available with a commercial and Medicare versions for each. Milliman starts by classifying diagnosis, pharmacy, and procedure codes into MARA categories, then applies relative weights to the categories for a given patient[53]. The MARA DxAdjuster output results in four risk domain scores and two composite scores.

1) Inpatient Risk Score
2) Outpatient Risk Score
3) Pharmacy Risk Score
4) Physician/ Other Risk Score
5) Emergency Department
6) (1+2+3) = Medical Risk Score (Composite)
7) (1+2+3+4+5) = Total Risk Score

Scores one through five are calculated by summing the cumulative risk weights within each of the MARA categories[53]. The categorical method provides deeper insight into risk drivers. The service categories also have varying predictability. One study by showed the following R^2 values by category: inpatient facility (6.0%), outpatient facility (20.1%), professional (20.2%), pharmaceutical (59.9%), total (28.8%) [57]. This level of granularity holds the potential to facilitate targeted interventions. The MARA performed best with the combined diagnosis and pharmacy prospective model with a R^2 of 20.7% uncensored, and a R^2 of 27.7% when censored at $250,000[14].

Centers for Medicare and Medicaid Services HHS-HCC Model v3

The Medicare Hierarchical Condition Category (HCC) system was created to risk-adjust capitation payments for Medicare Part C plans. These payments were originally calculated using geographic location and the fee-for-service expenditures[29]. In 2000, Medicare began adjusting 10% of capitation payments based on the Principle Inpatient Diagnostic Cost Group Model (PIP-DCG)[29]. This method neglected to adjust payments for patients who were not admitted to a hospital. This had the unintended effect of reducing capitation payments for patients who were being well managed in ambulatory settings and were thus not being admitted [29,58]. Medicare then evaluated several risk adjustment models and ultimately chose the HCC model due to its transparency, ability to be modified, and clinical coherence [29]. The HCC model development was funded by CMS and developed by RTI International, Boston University, and Harvard Medical School [29]. The model was developed using the following principles:

1. Diagnostic categories should be clinically meaningful
2. Diagnostic categories should predict medical (including drug) expenditures
3. Diagnostic categories that will affect payments should have adequate sample sizes to permit accurate and stable estimates of expenditures.
4. In creating an individual's clinical profile, hierarchies should be used to characterize the person's illness level within each disease process, while the effects of unrelated disease processes accumulate
5. The diagnostic classification should encourage specific coding.
6. The diagnostic classification should not reward coding proliferation

7. Providers should not be penalized for recording additional diagnoses (monotonicity)
8. The classification system should be internally consistent (transitive).
9. The diagnostic classification should assign all ICD-9-CM or ICD-10- CM codes (exhaustive classification)
10. Discretionary diagnostic categories should be excluded from payment models[29]

The HCC model uses the following steps:

1. Classify thousands of diagnosis codes exclusively into 804 Diagnostic Groups
2. Diagnostic Groups are further categorized into 189 Condition Categories
3. Patients are placed in the most severe category using a hierarchical ranking of the most severe category for a related disease.

Diagnosis data is collected through primary hospital inpatient data, secondary hospital inpatient data, hospital outpatient data, physician professional billing data, and non-physician clinical data[29]. During development, the HCC predictive power was improved by adding diagnosis data from home health (R2 from 11.15 to 11.65) and durable medical equipment claims (11.65 to 11.85) while adding data from laboratory and radiology claims made prediction less accurate due to the complexities of rule-out diagnosis process [29]. The final model retains 70 of the original 189 Condition Categories which are similar to ICD-9 major cost categories. They are comparatively less homogenous but the categories are similar both clinically and similar in cost [29].

Essential Differences Between Proposed Research and Prior Studies

The proposed research will expand upon existing research in a number of ways. Others have used diagnosis and claims grouping and classifying models to predict high future cost. This research will use features mapped across ontologies and machine learning to build models to predict outcomes of interest. The use of EDW ontology mappings as features in a machine learning classifier model in this area is novel. The research will incorporate clinical data from disparate EMR's which has not been included in any of the claims-based risk adjustment models[14]. The data aggregated by the researcher is unique in scope and form. Aggregating clinical data from across a health system and joining this data with claims data creates a potential for a rich dataset. The novel approach using unique data assets benefit both healthcare and informatics fields. Additionally, other organizations with similar EDW's could deploy this method to predict outcomes.

Previous Work

The researcher authored a previous paper outlining a strategy for risk stratifying the HMH CIN and CPC+ patient populations. The method used in this strategy involved using a regression-based proprietary risk model, MARA, for patient for whom the organization had adjudicated claims. For all patients who had no adjudicated claims, the Charlson co-morbidity index was used to stratify patients. Six risk strata for Medicare and commercial populations determined by Milliman were used along with the Charlson

index scores to combine the three models into one. The resulting model placed patients into three categories, low, moderate, and high. Table IV-1 shows the blended model.

Table 5 MARA and Charlson Comorbidity Index Blended Model[7]

MARA	Commercial	Medicare	Charlson	**FINAL**
Very High	High	High	High	**HIGH**
High	High	High		
Moderate High	Moderate	Moderate	Moderate	**MODERATE**
Moderate	Moderate	Moderate		
Low	Low	Low	Low	**LOW**
Very Low	Low	Low		

This method was successfully used to stratify the target population for care management activities.

Population Characteristics

All HMH ACO Patients 191,000 limited to all in both 2016 and 2017 resulted in a population of 76,345. The difference between these numbers is driven by non-continuous coverage over the 24 continuous months in 2016 and 2017. In real-world applications the researcher has observed significant movement in beneficiary coverage over time and across products. The study population is 55% female and 45% male. The age distribution can be seen in figure 6 and race distribution in table 6. Additional population characteristics can be seen in the data section (5.3 Data).

Figure 6 Study Population Age Distribution

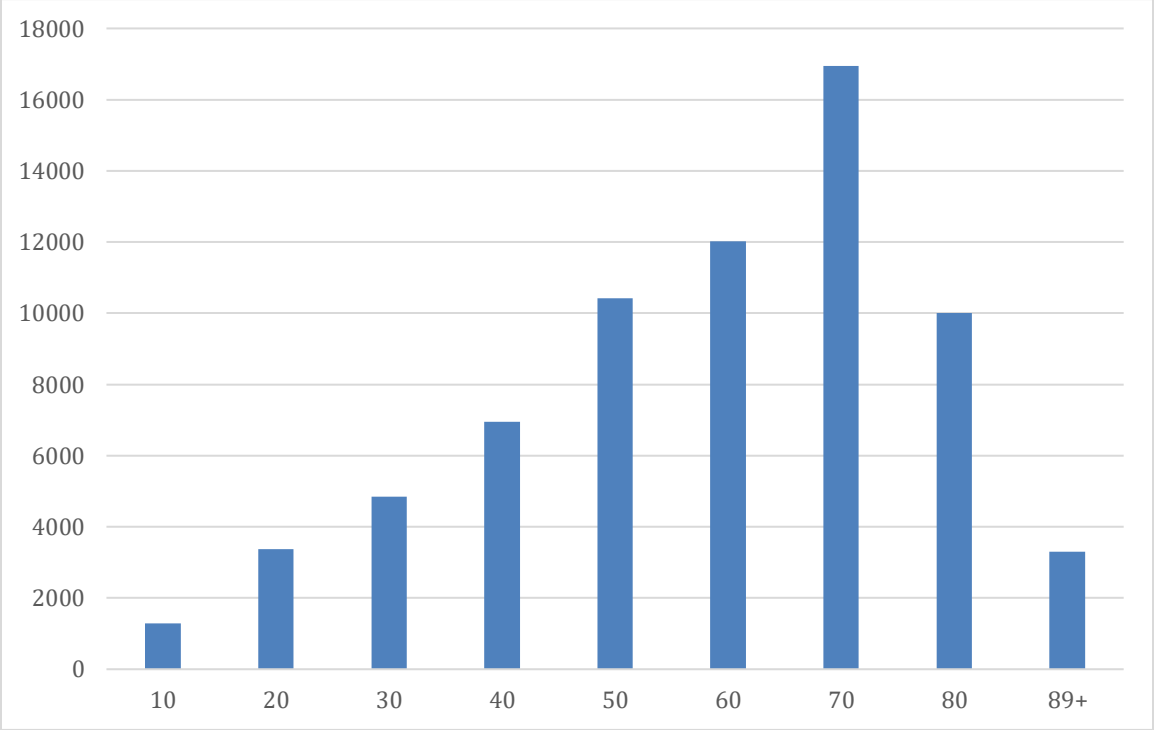


Table 6 Study Population Race Distribution

Race	Count
CAUCASIAN	33,274
UNKNOWN	32,585
BLACK OR AFRICAN AMERICAN	2,614
ASIAN	1217
HISPANIC	239
OTHER	206
OTHER PACIFIC ISLAND	94
AMERICAN INDIAN OR ALASKA NATIVE	87
MULTIRACIAL	29

Chapter III

METHODS

The research used de-identified data from the Hackensack Meridian Health Population Health Enterprise Data Warehouse.

Data Acquisition

The Cerner HealtheIntent Platform is an information technology platform that offers a suite of solutions. At it's core, HealtheIntent is a cloud-based platform designed to integrate vendor-agnostic disparate data sources using Hadoop technology. Data is normalized and standardized into a population record that can be subsequently used for population health management and clinical care. Cerner maintains ontology mappings as a service and has a team of clinical terminologists constantly updating ontology mappings for both standard ontologies and proprietary codes from source systems.

Patient Matching

Matching patients from disparate data sources is a significant challenge in any enterprise-wide health analytics framework since there is no single universally applicable

and unique person key shared across the sources. To mitigate the issues of duplicate patient entries leading to incorrect analyses and reports, many organizations come up with novel 'matching' solutions (normally subject to Intellectual Property Rights), devised using an admixture of various statistical and algorithmic techniques. In this research we used an enterprise master patient index (EMPI) solution, as employed by Cerner (through the HealtheIntent platform). This EMPI solution uses a *greedy* algorithm along with string comparison measurement (*Levenshtein Distance)*, to compare name, date of birth, demographic data, and alias identifiers resulting in a weighted match probability score ranging from zero to one[59,60]. This score is then used to categorize patient entry matches into three categories, *no match*, *manual review*, and *automatic match*. The entries in the manual review category go before human review where match decisions are made while the other two are processed appropriately. This approach has been shown to reduce the comparison burden dramatically and results in efficient, accurate patient matching across the platform (compared to 'brute-force' techniques) in a time-parsimonious manner[60].

Data

The HMH instance of HealtheIntent has a variety of data types being loaded to the platform.

Conditions

The conditions are diagnoses mapped from all data sources. The ontology mappings result in 583 distinct conditions and 990,547 records for the research population.

Results

The data type "Results" refer to lab and testing results. The dataset contains 3,537 distinct result types and 9,746,522 records.

Procedure

Procedures are medical procedures often coded using CPT from raw data sources. The ontology mappings resulted in 77 distinct procedures and 217,866 records.

Medication

Medications from all connected sources mapped through the ontology mappings resulted in 1,708 distinct medications and 124,241 records.

Cost

The total cost for the population in 2016 was $ 558,947,972. There were 5,767 patients had zero cost in 2016. The mean cost was $7946. The median cost in 2016 was $2,187.

Table 7 Cost Distribution

2016 Percentile	Cost in $
.05	0
.10	98
0.25	632
0.50	2,187
0.75	6,138
0.90	18,454
0.95	35,490

2017 Percentile	Cost in $
0.5	159
0.10	304
0.25	844
0.50	2,617
0.75	7,517
0.90	24,620
0.95	45,015

Total paid in 2017 was $678,431,927 with a $9,644.35 per member per year cost, and $803.70 per member per month cost.

Demographics

Demographic data from payer sources are considered first choice in a hierarchy followed by EMR and other data sources within the EDW. When there is a discrepancy in demographic data the data from the payer is used.

Claims

Claims data typically comes in three file types, Enrollment, Medical Claims, and Pharmacy Claims. Some payers provide additional files such as lab data files and utilization and quality reports. Claims files are received for roughly 300,000 patients at

HMH. These claims are loaded to the HealtheIntent platform and are then processed by Milliman to calculate the Milliman Advanced Risk Adjusters (MARA) risk scores. The claims are also processed by 3M™ for All Patient Refined DRG (APR DRG) Classification, Clinical Risk Grouping (CRG) Classification, and Enhanced Ambulatory Patient Grouping (EAPG).

Electronic Medical Record

Hospital-based EMR

The hospital EMR's contributing data to the platform include the following hospitals;

Table 8 Hospitals Contributing Data to EDW

Hospital	EMR	Billing
1) Hackensack University Medical Center	Epic	Epic
2) Jersey Shore University Medical Center	Soarian	Invision
3) Ocean Medical Center	Soarian	Invision
4) Riverview Medical Center	Soarian	Invision
5) Bayshore Medical Center	Soarian	Invision
6) Southern Ocean Medical Center	Soarian	Invision
7) Raritan Bay Medical Center – Perth Amboy	Soarian	Invision
8) Raritan Bay Medical Center – Old Bridge	Soarian	Invision
9) Palisades Medical Center	Soarian	Invision

The following data areas are captured through Hospital EMR and billing systems.

1. Clinical
2. Billing
3. Emergency Department
4. Hospital-based Ambulatory

Ambulatory

Epic Ambulatory

The employed physicians affiliated with Hackensack University medical Center have used the Epic EMR for ambulatory care delivery for a number of years. Outpatient clinical documentation and billing is documented through Epic which s extracted in flat files then loaded to HealtheIntent.

GE Centricity EMR

The employed physicians affiliated with Jersey Shore University Medical Center, Ocean Medical Center, Riverview Medical Center, Bayshore Medical Center, and Southern Ocean Medical Center used GE Centricity until the end of 2017 at which time they changed to Epic. Both the data from the GE Centricity EMR and Billing systems have been loaded to the HealtheIntent platform.

Home Care

The HMH home care agency uses Cerner Home Works in all regions where it provides home care services. The data from this system is extracted from Cerner Home Works and loaded to HealtheIntent.

Long-Term Care and Rehab

Hackensack Meridian Health owns the following long-term care and rehabilitation facilities:

1) Meridian Nursing & Rehab at Brick
2) Meridian Nursing & Rehab at Bayshore
3) Meridian Nursing & Rehab at Manor by the Sea
4) Meridian Nursing & Rehab at Shrewsbury
5) Meridian Willows Assisted Living Community

These facilities use the Sigmacare EMR. Data from these facilities is extracted using the Consolidated Clinical Document Architecture (CCDA) and loaded to HealtheIntent.

Health Information Exchange

Health Information Exchanges (HIE) facilitate healthcare data exchange between providers in a region. The purpose of sharing the data is to increase continuity and communication or information sharing between providers and healthcare facilities to reduce waste and improve coordination. Most HIE interfaces transmit Amission Discharge Transfer (ADT) information as well as CCDA's. The state of New Jersey has had a number of HIE's but none that have spanned the entire state or into neighboring metropolitan areas across state lines into New York or Philadelphia.

Jersey Health Connect

Jersey Health Connect is the largest Health Information Exchange in the state of New Jersey with over 30 of the states hospitals, thousands of physicians and medical offices, multiple homecare agencies, and over 140 long-term care facilities. HMH provides the HIE technology vendor a roster of commercial and Medicare ACO patients and a data extract from the HIE is then loaded to HealtheIntent. This data is limited to patient demographics, medications, immunizations, allergies, encounters, and results. While the data is limited in depth, it is broad in reach spanning all of the contributing systems contributing data to the HIE.

Patient Ping

Patient Ping is a patient tracking solution that uses HL7 Admission Transfer Discharge (ADT) interfaces to track patients across care venues. Patient Ping has a strong presence in the post-acute care space as will as inpatient facilities. The Patient Ping network spans across state lines and into other regions. Coverage is not contiguous however any ADT transaction from across all interfaced systems can be used to track patients. The tool allows teams to track patients when they leave their venue of care and can be used to improve coordination, utilization, and quality. Transactions collected through Patient Ping are extracted and loaded to HealtheIntent.

Lab Data

Two of the largest commercial lab services providers share lab data for HMH mutual patients. Both of these lab vendors receive patient and provider roster and in turn

send lab data which is loaded to the platform. This data supplements the lab data collected through all of the contributing EMR systems previously mentioned.

LabCorp

LabCorp is one of the largest laboratory diagnostic services companies in the state of New Jersey. Results from LabCorp are interfaced to many of the EMR systems feeding into the enterprise data warehouse. These interfaced results start as an order from the EMR system, the specimen is collected and analyzed, results are interfaced back to the EMR, the results are extracted transformed and loaded to the enterprise data warehouse and then ultimately become part of the patient's population record. In addition to this

Quest Diagnostics

Similar to the process used with LabCorp, Quest diagnostics shares diagnostic results for HMH ACO patients via flat file extract. These files are similarly loaded to the platform, standardized, normalized, and matched to patients.

Data Cleansing

Data was cleansed by checking ranges, deleting impossible occurrences (i.e. gender/age-based etc.), remove indecipherable data, remove non-conforming data, identify missing values [19]. Missing values were replaced with '0', to facilitate the Random Forest model processing. The HCUP Quality Control Procedures were emulated by reviewing numeric value means, missing and non-missing frequencies, and minimum

and maximum values[61]. The maximum number of categories per variable was limited to 32 to allow processing of the RandomForest R package.

Structure of the Data

The data was structured into a format with a de-identified ID, features, and outcomes as outlined in Table 9.

Table 9 Data Structure for Random Forest

Category	**Column**	**Data Type**
Key	MLRSID	Identity
Features	Birth Year	Integer
	Zip Category	Category
	Race	Category
	Gender	Category
	Condition 1	Binary
	Condition …	Binary
	Condition 583	Binary
	Procedure 1	Binary
	Procedure …	Binary
	Procedure 77	Binary
	Medication 1	Binary
	Medication …	Binary
	Medication 1,708	Binary
	Result 1	Integer/Category
	Result…	Integer/Category
	Result 722	Integer/Category
	2016 Cost	Integer
	2016 Admission Count	Integer
	2016 ED Count	Integer
Outcomes	Admission	Binary
	Readmission	Binary
	Multiple ED Visits	Binary
	High Cost	Binary

Data Analysis: Random Forest

Leo Breiman developed the Random Forest model, a machine learning algorithm designed for both regression and classification problems[62]. This research focuses on the classification type problem. The algorithm is an ensemble method that combines many decision trees into one model. The decision trees are created using random bootstrap aggregation, or *bagging,* to create many small random samples using replacement. In addition to bagging, the model uses random feature selection at each tree[63,64]. The method builds trees using small samples of observations and features. The trees are not pruned as they would be using the traditional Classifier and Regression Tree (CART) methodology[62,65]. The combination of bagging, random feature selection, and unpruned trees results in classification trees that have low correlation and an ensemble with relatively low bias and variance[66].

Classification Decision Trees

The CART method creates a decision tree by minimizing the error in each node of a tree. For all variables, data is separated into two maximally homogenized groups at each split[67]. The final tree is completed by assembling an over-grown tree which is pruned to an optimal size where size is equal to the number of groups[68]. Binary variables are grouped in their two classes, and variables with more than two categories are split using every possible combination of categories. With k equaling the number of variables, the number of possible spit combinations can be seen in Equation 1[68].

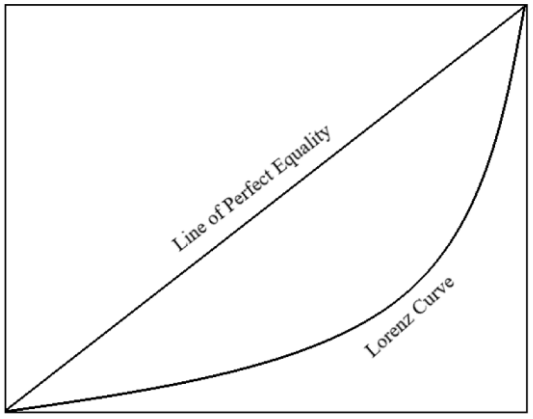
Equation 1

$$2^{k-1} - 1$$

Gini Measure of Impurity

The Random Forest model uses the Gini measure of impurity to determine the lowest level of impurity for each node split[69]. The Gini index is the area between the line of perfect equality and the Lorenz curve divided by the area below the line of perfect equality[70].

Figure 7 Visualizing the Gini Index



The majority vote for mtry samples is selected at each split in a given tree.

Many Features

This algorithm has been shown to be well suited for classification when there are many features in a data set[71]. This is due to the random sampling and random feature selection. The ensemble approach performs well when there are many features and sparse data.

Overfitting

The method has produced strong models without overfitting, a phenomenon that occurs when a decision model uses idiosyncrasies in the training data to the point where the model does not perform well on real data [71]. Each tree is grown independently and thus the resulting ensemble is robust against overfitting.

Managing Noise in Data

The Random Forest model has also shown it is robust against noise within data [62,71]. This is due in part to the bagging method used by the Random Forest model. Bagging has been shown to be the most robust against noise when compared to boosting and randomization[72]. The use of an ensemble of trees is also protective against noise in the data.

Imbalanced Classes

A binary classifier's performance measurement can be misleading when classes are imbalanced. When an outcome has a relatively small number of instances, i.e. positive diagnosis of a specific disease, a model can predict all outcomes to be negative, and have a low error rate. The outcomes in this study are imbalanced. To address the issue of imbalanced classes, the models were trained with artificially balanced samples. The sample sizes were 300, 300 or 500, 500 depending on the outcome frequency. The sample sizes are reported in each models result section.

Feature Selection and Curation

Feature selection and curation is generally not required when using Random Forests. The model performs feature selection randomly then uses voting to determine node impurity splits. The model performs well with highly dimensional data and was selected for this research in part because of this characteristic[62,73].

Feature Multicollinearity

Random Forest models have shown strong predictive ability in the presence of highly correlated features[73]. Feature collinearity dose however pose a problem when trying to determine variable importance in the models. Breiman originally used Gini impurity reduction to rank variable importance, an approach which has been shown to be biased towards features with higher number of categorical values[73,74]. An approach using conditional variable importance and conditional inference trees may address this issue[73]. Feature importance within data types is not explored in this research.

Random Forest Model Performance

There are a number of ways Random Forest model performance can be evaluated. One method of measuring accuracy of a binary classifier is to use an Error Matrix, see Table 10. When a model predicts the binary outcome to be true and the actual outcome is true, it is termed a *true positive*. When the model predicts the outcome to be true and the actual outcome is false, it is termed a *false positive*. When a model predicts an outcome to be false and the actual outcome is false, it is termed *true negative*. Finally, when a

model predicts an outcome to be false, and the actual outcome is positive it is termed a *false negative*.

Table 10 Error Matrix

		Actual Class	
		Positive	Negative
Predicted Class	Positive	True Positive = TP	False Positive = FP
	Negative	False Negative = FN	True Negative = TN

Equation 2 Error Overall Accuracy

$$\text{Overall Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Equation 3 Precision

$$\text{Precision} = \frac{TP}{TP+FP}$$

Equation 4 Sensitivity, Recall, or True Positive Rate (TPR)

$$\text{Recall} = \frac{TP}{TP+FN}$$

Equation 5 False Positive Rate

$$\text{False Positive Rate} = \frac{FP}{TP+TN+FP+FN}$$

Equation 6 F_1 Score: Harmonic Mean of Precision and Sensitivity

$$F_1 \text{ Score} = \frac{2TP}{2TP+FP+FN}$$

These methods do not measure model performance well when classes are skewed[75,76]. These will be reported as a point of reference however the models will be evaluated using the Receiver Operating Characteristic (ROC) curve.

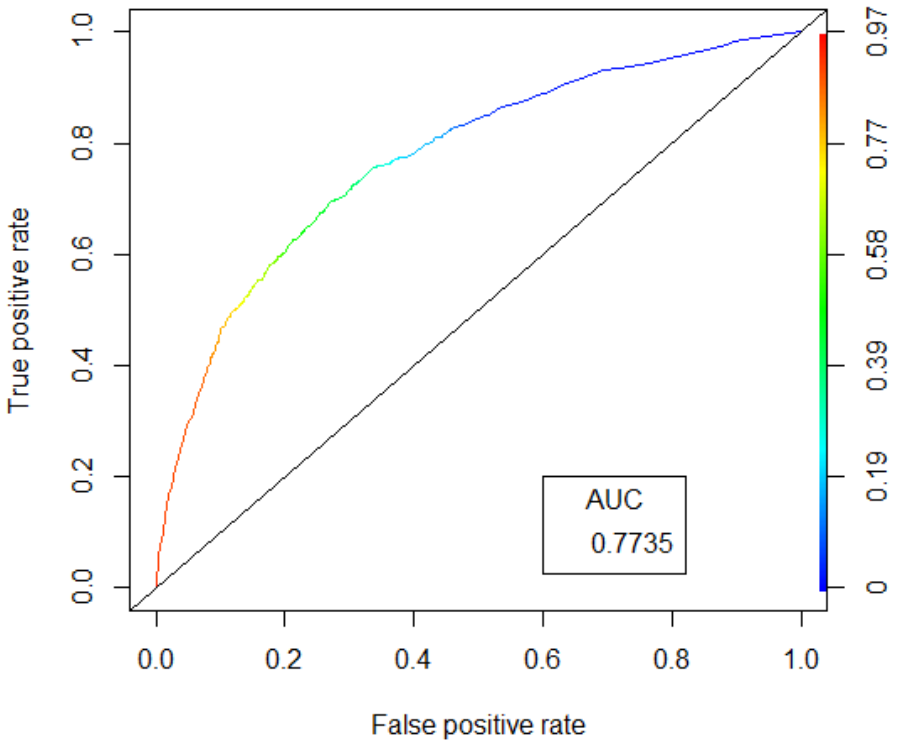
Receiver Operating Characteristic Curve (ROC)

The ROC curve has been recommended as the preferred method for evaluating binary classifiers especially when there are imbalanced classes and disproportionate cost for misclassification such as the classes in this study[75]. It has also been identified as the preferred instrument in measuring accuracy in disease management programs[77]. The ROC curve is unaffected by imbalanced classes and will appear the same with or without class imbalance. The ROC curve is a two-dimensional graph that is constructed by plotting the true positive rate on the Y axis and the false positive rate on the X axis for each cutoff point in class probabilities[75]. The line with 0 intercept and slope of 1 represents an AUC of 0.5. Points along this line represent a random classifier[75]. An AUC of 1 represents a perfect classifier and an AUC less than 0.5 represents a model that is worse than chance and if negated, would have a positive predictive AUC[75]. The default cutoff point used in the randomForest package in R is 0.50. In Figure 8 you will see the cutoff values on the right side of the plot along the y axis. If the cutoff value were to be changed to 0.75 the true positive rate would drop to roughly 0.55 and the false positive rate would drop to roughly 0.15. This is one of the key benefits of the ROC

curve, it allows the visualization of all cutoff points for a classifier and a single number can be used to measure the curve and compare to other classifiers. The area under the curve in Figure 8 is 0.7735. Points along the curve that are closest to the top-left of the plot can be considered superior cutoff points as true positives are higher and false positives are lower as you approach the point (0,1).

The outcome of the models are reported using the testing dataset and the Area under the curve (AUC) of the Receiver Operating Characteristic (ROC) for the four classification outcomes of interest across all models. The OOB estimate of error can also be used as a measure of model performance and is calculated using unsampled data from the train data set[62]. This result is also displayed for each model.

Figure 8 ROC Plot



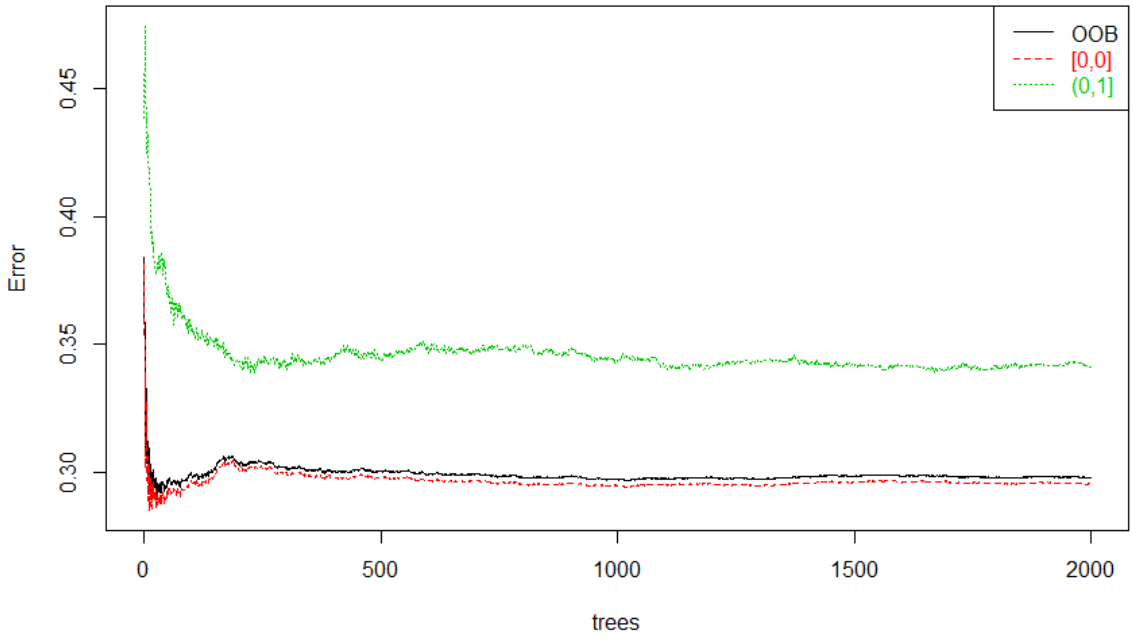
Precision Recall Curve

The Precision Recall (P/R) curve plots the positive predictive value (PPV) or Precision on the y axis and the TPR or Recall on the x axis. The Precision Recall plot visualizes information across the entire model similar to the ROC curve[76]. Precision and recall are useful metrics for imbalanced datasets because they focus on the positive outcome. The baseline for a given precision recall curve is the determined by the ratio of positives divided by total. Unlike ROC curves, P/R curves are impacted by class imbalance. Along with visualizing the P/R curve, the F_1 score can be used as a single value metric for precision and recall. The calculation for this measure can be seen in Equation 6. The F_1 score punishes the lower of the two values, either precision or recall.

Tuning Random Forest

There are only two parameters to consider when tuning a Random Forest. The number of trees, and the number of features sampled in those trees. The models are relatively insensitive to these tuning parameters. The default number of trees used in the R RandomForest package is 500. Plotting the error rate can help determine the optimal number of trees for the forest for computational efficiency. In Figure 8 you can see there is not significant gain in performance after 250 trees. The default of 500 trees was used in all models after observing no dramatic gain in performance. The RandomForest package in R defaults the number of variables parameter to the square root of the total number of features in each data set. This default was used for each model.

Figure 9 Error Rates with Two Thousand Trees



Partitioning Data

The method requires partitioning data into training and testing sets.

The bagging method resamples from the training dataset with replacement for each decision tree [71]. The models were run using R version 3.5.0 and the RandomForest package.

The research data set was be split into the following partitions using a seed value of 42:

Training: 70%

Testing: 30%

The training dataset was used to train each of the Random Forest models. The performance of the models was then measured using the reserved test data set. Except for

one measure, the Out-of-bag estimate of error which uses OOB samples from the training data set to estimate the error rate.

Outcome Distributions

The binary outcomes of interest in the research data set are significantly skewed as seen in Table 18. While training the models samples were balanced to 500 (negative) and 500 (positive) for all outcomes except Multi-ED which was balanced at 300, 300 due to the lower number of available positive outcomes in the training data set for this measure. This balancing of the training samples allows the Random Forest model to consider the positive outcomes appropriately for this research.

Table 11 Outcome Distributions Full Dataset

Outcome	No	Yes
High Cost	64,211	6,134
Admission	68,372	1,973
Readmission	69,840	505
Multi-ED	66,710	3,635

Chapter IV

RESULTS

The models were run with different data types. The data types used in each model are represented as follows:

U = demographics and utilization

C = conditions

P = procedures

M = medications

R = results

Models are labeled first with the target outcome then the abbreviation for the feature types used as inputs in the model. For example, High Cost (U, C, P) indicates the outcome was high cost and the model features included utilization and demographics, conditions, and procedures. Other feature types where excluded in this model.

Reporting Model Results

The tuning parameters for each model are listed as:

- ntree = number of trees

- mtry = number of features evaluated at each split in the trees
- Sample size = 500 or 300 depending on the outcome.

Plots were generated using the R ROCR package

The test data set performance of each model is reported using:

a) Out-of-bag estimate of error

b) Error Matrix of counts

c) Error Matrix of proportions

d) Overall accuracy

e) Precision

f) Recall

g) Specificity

h) F_1 Harmonic mean of precision and sensitivity

i) ROC Curve/AUC

j) Precision/ Recall Curve

Figure 10 Model Comparison

AUC				
	High Cost	Admission	Readmission	Multi-ED
U	0.80	0.80	0.84	0.71
C	0.78	0.77	0.85	0.73
P	0.73	0.64	0.67	0.69
M	0.60	0.53	0.51	0.58
R	0.64	0.60	0.57	0.63
UC	0.81	0.80	0.84	0.74
UCP	0.81	0.74	**0.86**	0.74
UCPM	**0.81**	**0.80**	0.85	**0.74**
UCPMR	0.79	0.79	0.83	0.73

(a)

Precision				
	High Cost	Admission	Readmission	Multi-ED
U	0.2007	0.0732	0.0303	0.0965
C	0.1958	0.0676	0.0302	0.1033
P	0.1737	0.0483	0.0177	0.1029
M	0.2166	0.0405	0.0113	0.1211
R	0.1591	0.0436	0.0130	0.0948
UC	0.2108	0.0773	0.0318	0.1046
UCP	0.2079	0.0754	0.0301	0.1035
UCPM	0.2117	0.0780	0.0340	0.1085
UCPMR	0.1930	0.0688	0.0283	0.1061

(b)

Recall				
	High Cost	Admission	Readmission	Multi-ED
U	0.7171	0.6775	0.7419	0.6432
C	0.7115	0.6537	0.6859	0.6578
P	0.6428	0.4949	0.4839	0.5703
M	0.2625	0.1164	0.0839	0.2345
R	0.4892	0.5060	0.3576	0.4780
UC	0.7395	0.6935	0.7308	0.6807
UCP	0.7514	0.7397	0.7355	0.7099
UCPM	0.7379	0.6901	0.7161	0.6979
UCPMR	0.7435	0.7163	0.6821	0.6676

(c)

Overall				
	High Cost	Admission	Readmission	Multi-ED
U	0.7298	0.7531	0.8239	0.6689
C	0.7068	0.7374	0.8347	0.6857
P	0.7060	0.7160	0.7989	0.7194
M	0.8548	0.8548	0.9393	0.8719
R	0.7338	0.6817	0.8013	0.7365
UC	0.7235	0.7626	0.8335	0.6824
UCP	0.7158	0.5095	0.8240	0.6657
UCPM	0.7252	0.7658	0.8483	0.7028
UCPMR	0.7106	0.7264	0.8301	0.6915

(d)

F1				
	High Cost	Admission	Readmission	Multi-ED
U	0.3136	0.1322	0.0583	0.1679
C	0.3071	0.1225	0.0578	0.1786
P	0.2735	0.0879	0.0341	0.1743
M	0.2374	0.2374	0.0199	0.1597
R	0.2402	0.0803	0.0251	0.1582
UC	0.3281	0.1392	0.0609	0.1813
UCP	0.3256	0.1369	0.0578	0.1807
UCPM	0.3291	0.1402	0.0648	0.1878
UCPMR	0.3064	0.1255	0.0544	0.1831

(e)

Specificity				
	High Cost	Admission	Readmission	Multi-ED
U	0.7310	0.7552	0.8245	0.6703
C	0.7064	0.7398	0.8358	0.6872
P	0.7119	0.7223	0.8012	0.7276
M	0.9106	0.9106	0.9456	0.9068
R	0.7568	0.6866	0.8045	0.7506
UC	0.7219	0.7645	0.8343	0.6825
UCP	0.7123	0.4968	0.8247	0.6633
UCPM	0.7239	0.7680	0.8493	0.7030
UCPMR	0.7075	0.7267	0.8312	0.6928

(f)

OOB Estimate				
	High Cost	Admission	Readmission	Multi-ED
U	26.84	24.93	17.61	33.33
C	29.10	26.28	16.73	31.41
P	29.64	28.32	19.98	28.30
M	14.23	14.23	5.99	12.58
R	26.48	31.90	19.62	26.54
UC	27.89	23.65	16.67	31.86
UCP	28.62	25.82	17.68	33.63
UCPM	27.62	23.42	15.54	29.99
UCPMR	29.24	27.03	16.94	31.10

(g)

High Cost (U)

ntree = 500
mtry = 2
Sample Size = 500, 500
OOB Estimate of Error = 26.84%

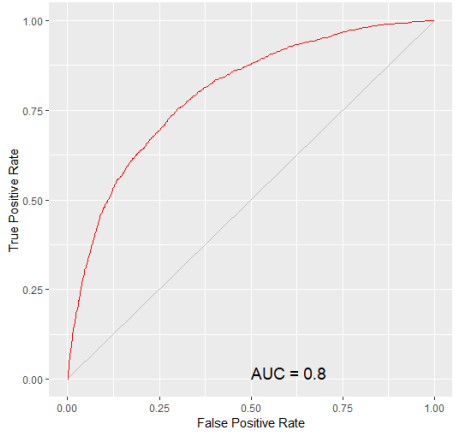
Table 12 High Cost (U) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	14,098	5,189	26.9
1	514	1,303	28.3

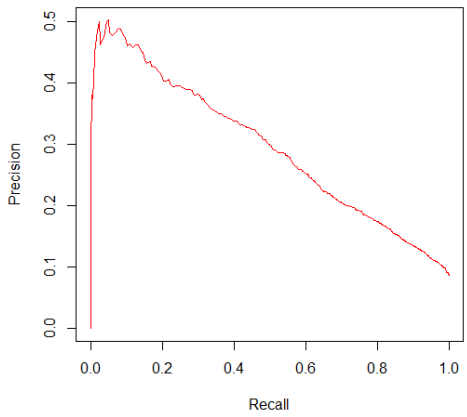
Proportions			
	Predicted		
Actual	0	1	Error
0	66.8	24.6	26.9
1	2.4	6.2	28.3

Overall Accuracy	0.7298
Precision	0.2007
Recall or Sensitivity	0.7171
Specificity TNR	0.7310
Harmonic Mean of Precision and Sensitivity	0.3136
AUC	0.80

Figure 11 High Cost (U) Test Curves



(a) ROC Curve



(b) P/R Curve

High Cost (C)

ntree = 500
mtry = 22
Sample Size = 500, 500
OOB Estimate of Error = 29.1%

Table 13 High Cost (C) Test Error Matrix

Counts

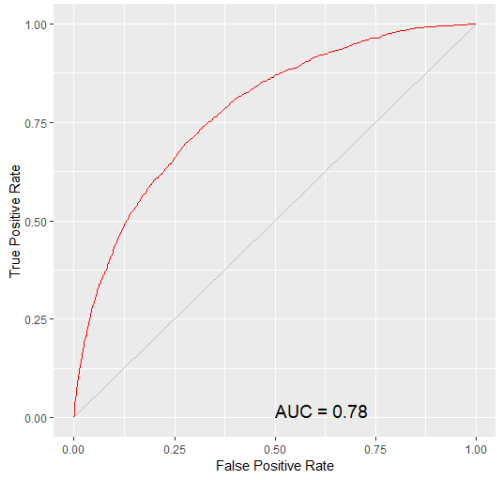
	Predicted		
Actual	0	1	Error
0	13,546	5,631	29.4
1	556	1,371	28.9

Proportions

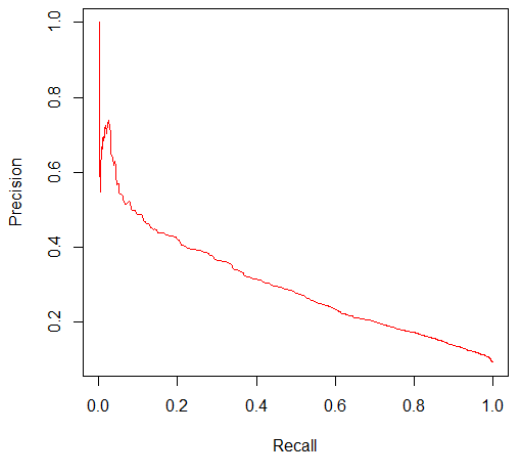
	Predicted		
Actual	0	1	Error
0	64.2	26.7	29.4
1	2.6	6.5	28.9

Overall Accuracy	0.7068
Precision	0.1958
Recall or Sensitivity	0.7115
Specificity TNR	0.7064
Harmonic Mean of Precision and Sensitivity	0.3071
AUC	0.78

Figure 12 High Cost (C) Test Curves



(a) ROC



(b) P/R

High Cost (P)

ntree = 500
mtry = 8
Sample Size = 500, 500
OOB Estimate of Error = 29.64%

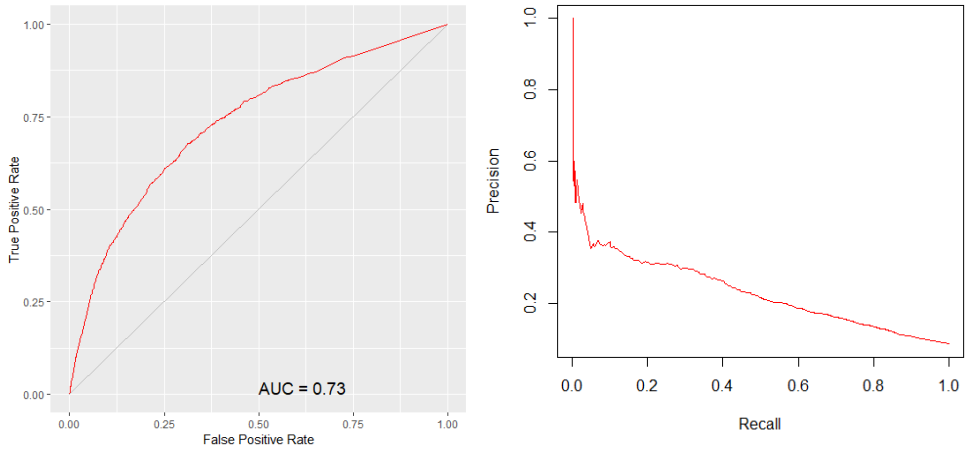
Table 14 High Cost (P) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,731	5,556	28.8
1	649	1,168	35.7

Proportions			
	Predicted		
Actual	0	1	Error
0	65.1	26.3	28.8
1	3.1	5.5	35.7

Overall Accuracy	0.7060
Precision	0.1737
Recall or Sensitivity	0.6428
Specificity TNR	0.7119
Harmonic Mean of Precision and Sensitivity	0.2735
AUC	0.73

Figure 13 High Cost (P) Test Curves



(a) ROC (b) P/R

High Cost (M)

ntree = 500
mtry = 26
Sample Size = 500, 500
OOB Estimate of Error = 14.23%

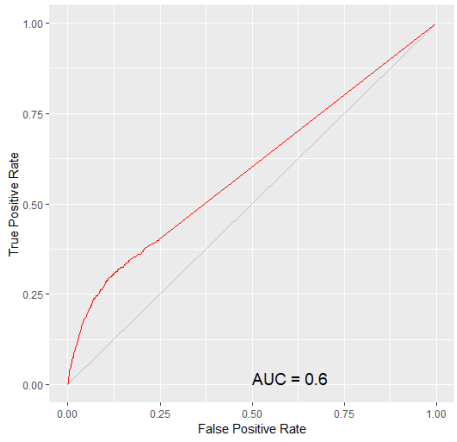
Table 15 High Cost (M) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	17,562	1,725	8.9
1	1340	477	73.9

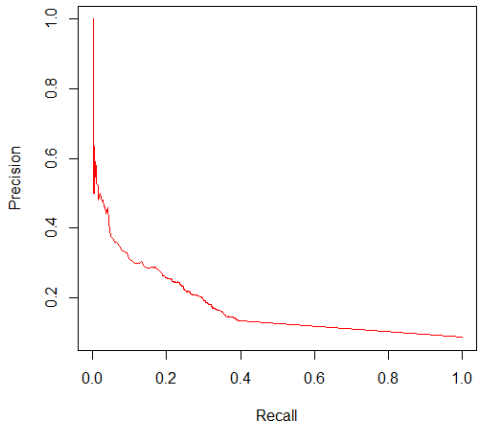
Proportions			
	Predicted		
Actual	0	1	Error
0	6.3	2.3	8.9
1	6.3	2.3	73.7

Overall Accuracy	0.8548
Precision	0.2166
Recall or Sensitivity	0.2625
Specificity TNR	0.9106
Harmonic Mean of Precision and Sensitivity	0.2374
AUC	0.60

Figure 14 High Cost (M) Test Curves



(a) ROC



(b) P/R

High Cost (R)

ntree = 500
mtry = 24
Sample Size = 500, 500
OOB Estimate of Error = 26.48%

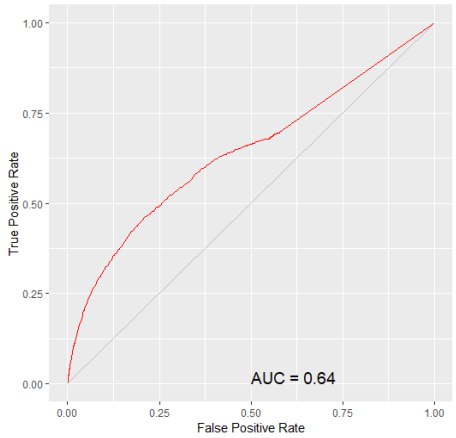
Table 16 High Cost (R) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	14,589	4,687	24.3
1	926	887	51.1

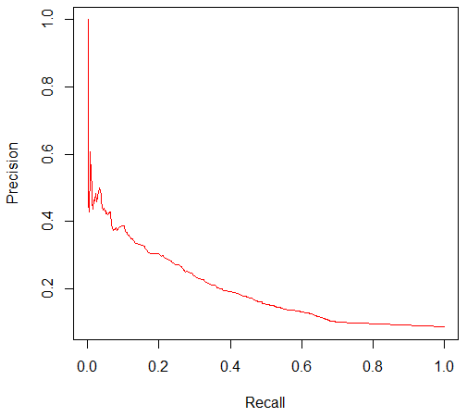
Proportions			
	Predicted		
Actual	0	1	Error
0	69.2	22.2	24.3
1	4.4	4.2	51.1

Overall Accuracy	0.7338
Precision	0.1591
Recall or Sensitivity	0.4892
Specificity TNR	0.7568
Harmonic Mean of Precision and Sensitivity	0.2402
AUC	0.64

Figure 15 High Cost (R) Test Curves



(a) ROC



(b) P/R

High Cost (U, C)

ntree = 500
mtry = 22
Sample Size = 500, 500
OOB Estimate of Error = 27.89%

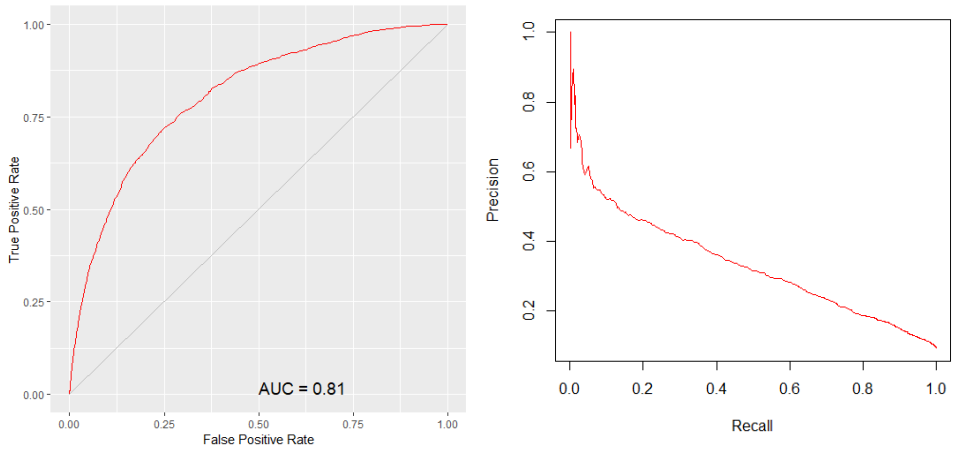
Table 17 High Cost (U,C) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,843	5,334	27.8
1	502	1,425	26.1

Proportions			
	Predicted		
Actual	0	1	Error
0	65.6	25.3	27.8
1	2.4	6.8	26.1

Overall Accuracy	0.7235
Precision	0.2108
Recall or Sensitivity	0.7395
Specificity TNR	0.7219
Harmonic Mean of Precision and Sensitivity	0.3281
AUC	0.81

Figure 16 High Cost (U,C) Test Curves



(a) ROC (b) P/R

High Cost (U, C, P)

ntree = 500
mtry = 24
Sample Size = 500, 500
OOB Estimate of Error = 28.62%

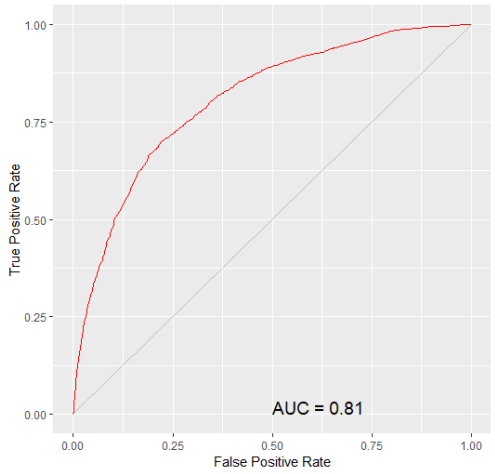
Table 18 High Cost (U,C,P) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,659	5,518	28.8
1	479	1,448	24.9

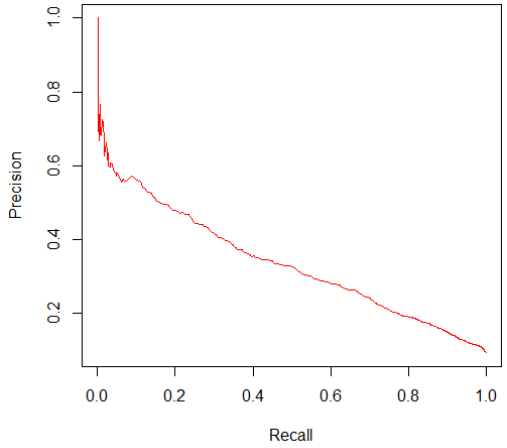
Proportions			
	Predicted		
Actual	0	1	Error
0	64.7	26.1	28.8
1	2.3	6.9	24.9

Overall Accuracy	0.7158
Precision	0.2079
Recall or Sensitivity	0.7514
Specificity TNR	0.7123
Harmonic Mean of Precision and Sensitivity	0.3256
AUC	0.81

Figure 17 High Cost (U,C,P) Test Curves



(a) ROC



(b) P/R

High Cost (U, C, P, M)

ntree = 500
mtry = 36
Sample Size = 500, 500
OOB Estimate of Error = 27.62%

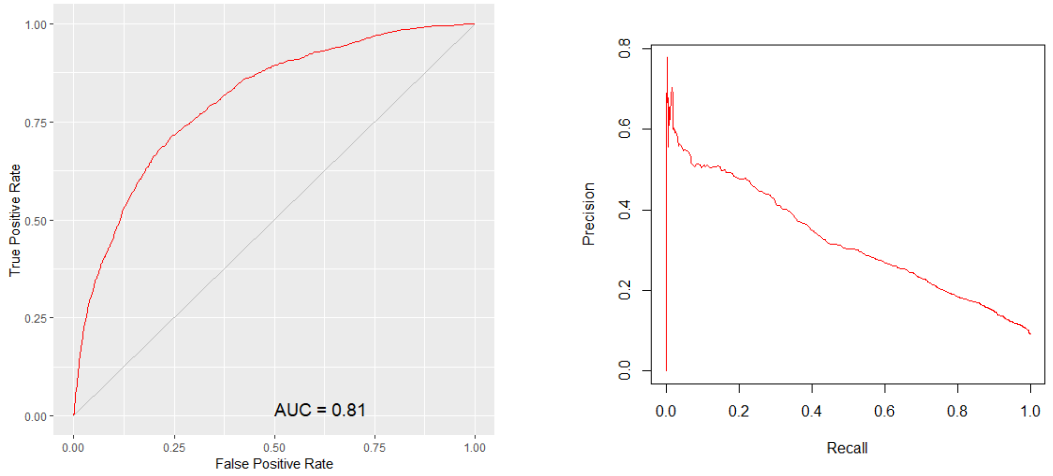
Table 19 High Cost (U,C,P,M) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,883	5,294	27.6
1	505	1,422	26.2

Proportions			
	Predicted		
Actual	0	1	Error
0	65.8	25.1	27.6
1	2.4	6.7	26.2

Overall Accuracy	0.7252
Precision	0.2117
Recall or Sensitivity	0.7379
Specificity TNR	0.7239
Harmonic Mean of Precision and Sensitivity	0.3291
AUC	0.81

Figure 18 High Cost (U,C,P,M) Test Curves



(a) ROC (b) P/R

High Cost (U, C, P, M, R)

ntree = 500
mtry = 43
Sample Size = 500, 500
OOB Estimate of Error = 29.24%

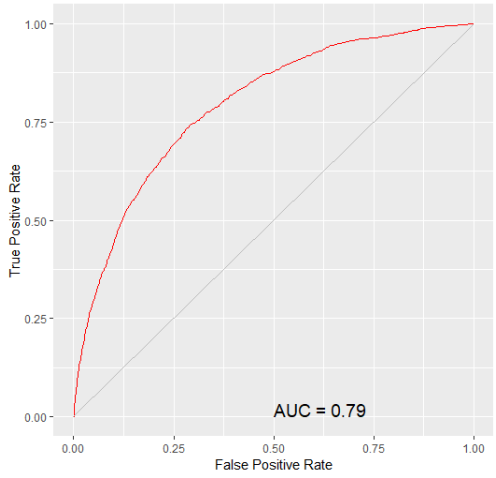
Table 20 High Cost (U, C, P, M, R) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,638	5,638	29.2
1	465	1,348	25.6

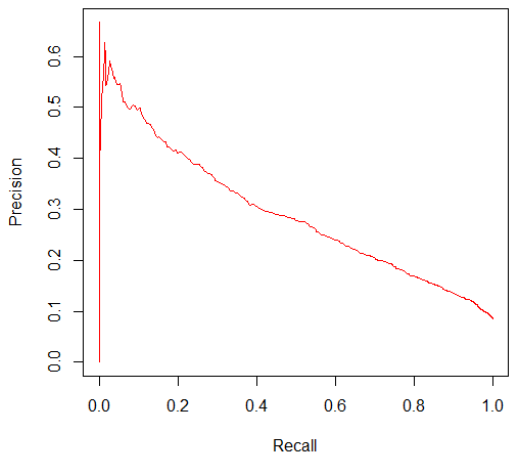
Proportions			
	Predicted		
Actual	0	1	Error
0	64.7	26.7	29.2
1	2.2	6.4	25.6

Overall Accuracy	0.7106
Precision	0.1930
Recall or Sensitivity	0.7435
Specificity TNR	0.7075
Harmonic Mean of Precision and Sensitivity	0.3064
AUC	0.79

Figure 19 High Cost (U, C, P, M, R) Test Curves



(a) ROC



(b) P/R

Admission (U)

ntree = 500
mtry = 2
Sample Size = 500,500
OOB Estimate of Error = 24.93%

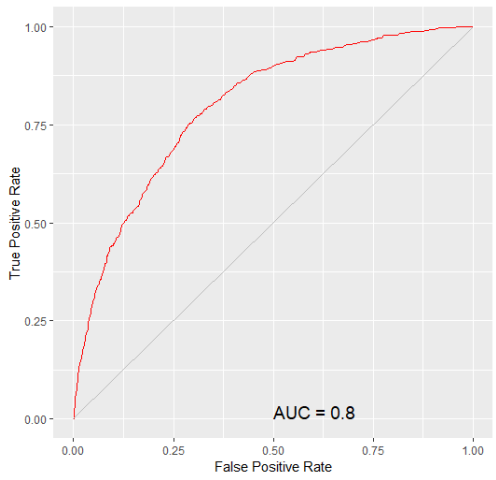
Table 21 Admission (U) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	15,497	5,023	24.5
1	178	406	30.5

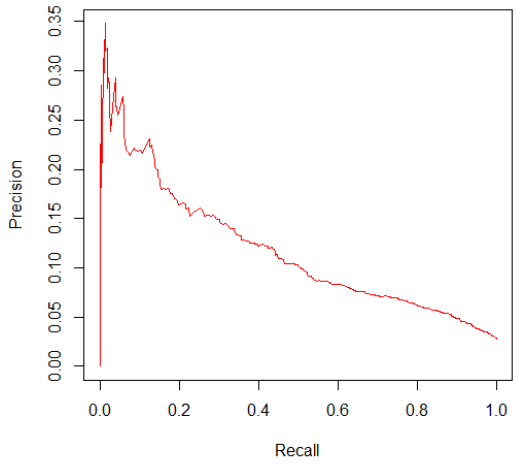
Proportions			
	Predicted		
Actual	0	1	Error
0	73.4	23.8	24.5
1	0.8	1.9	30.5

Overall Accuracy	0.7536
Precision	0.0748
Recall or Sensitivity	0.6952
Specificity TNR	0.7552
Harmonic Mean of Precision and Sensitivity	0.1350
AUC	0.80

Figure 20 Admission (U) Test Curves



(a) ROC



(b) P/R

Admission (C)

ntree = 500
mtry = 22
Sample Size = 500, 500
OOB Estimate of Error = 26.28%

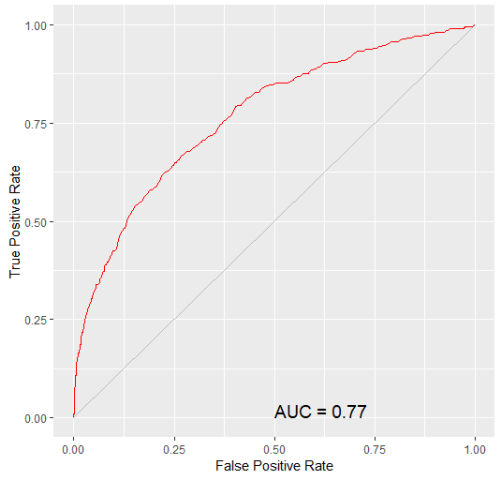
Table 22 Admission (C) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	15,175	5,337	26
1	205	387	34.6

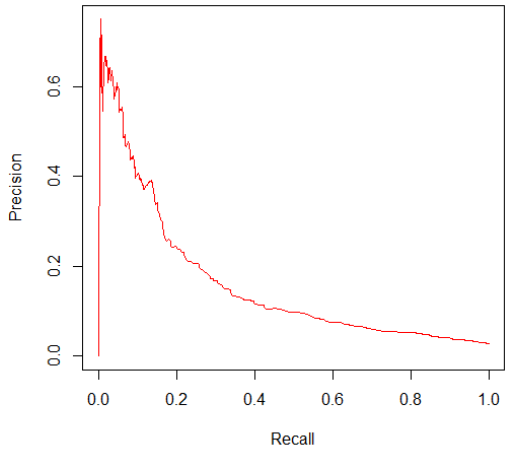
Proportions			
	Predicted		
Actual	0	1	Error
0	71.9	25.3	26
1	1	1.8	34.6

Overall Accuracy	0.7374
Precision	0.0676
Recall or Sensitivity	0.6537
Specificity TNR	0.7398
Harmonic Mean of Precision and Sensitivity	0.1225
AUC	0.77

Figure 21 Admission (C) Test Curves



(a) ROC



(b) P/R

Admission (P)

ntree = 500
mtry = 8
Sample Size = 500, 500
OOB Estimate of Error = 28.32%

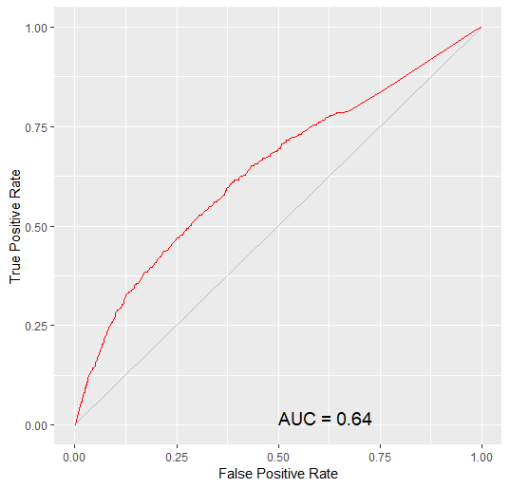
Table 23 Admission (P) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	14,821	5,699	27.8
1	295	289	50.5

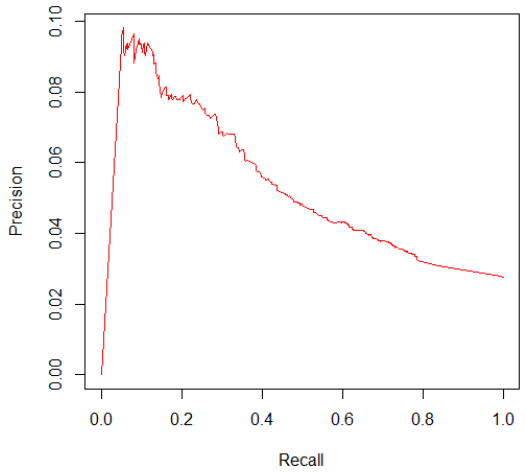
Proportions			
	Predicted		
Actual	0	1	Error
0	70.2	27	27.8
1	1.4	1.4	50.5

Overall Accuracy	0.7160
Precision	0.0483
Recall or Sensitivity	0.4949
Specificity TNR	0.7223
Harmonic Mean of Precision and Sensitivity	0.0879
AUC	0.64

Figure 22 Admission (P) Test Curves



(a) ROC



(b) P/R

Admission (M)

ntree = 500
mtry = 26
Sample Size = 500, 500
OOB Estimate of Error = 14.23%

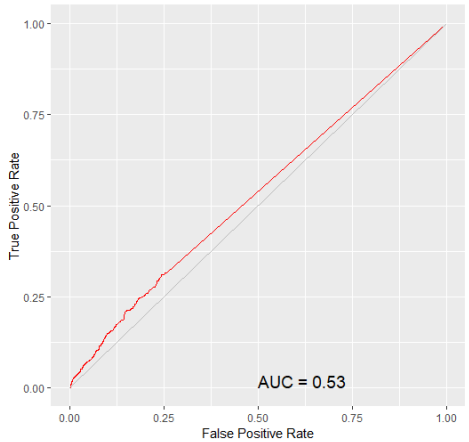
Table 24 Admission (M) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	18,910	1,610	7.8
1	516	68	88.4

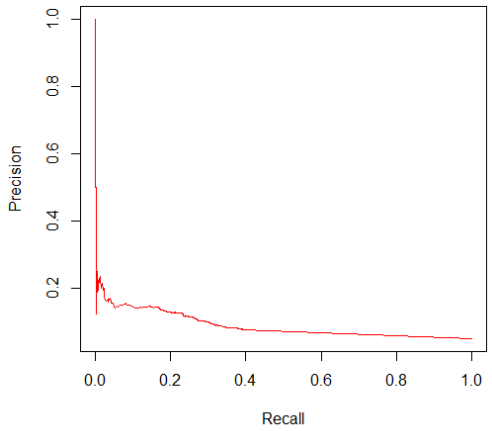
Proportions			
	Predicted		
Actual	0	1	Error
0	89.6	7.6	7.6
1	516	68	88.4

Overall Accuracy	0.8993
Precision	0.0405
Recall or Sensitivity	0.1164
Specificity TNR	0.9215
Harmonic Mean of Precision and Sensitivity	0.0601
AUC	0.53

Figure 23 Admission (M) Test Curves



(a) ROC



(b) P/R

Admission (R)

ntree = 500
mtry = 24
Sample Size = 500, 500
OOB Estimate of Error = 31.9%

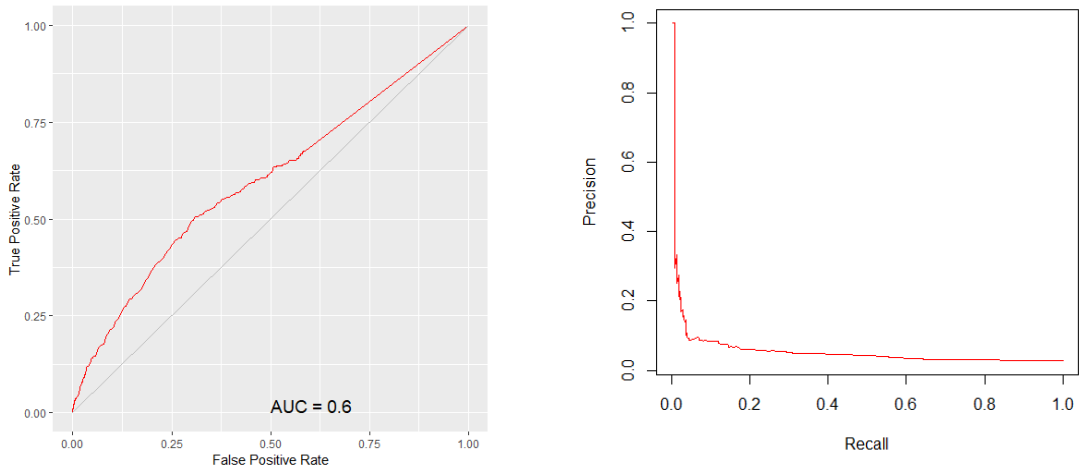
Table 25 Admission (R) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	14,083	6,428	31.3
1	286	293	49.3

Proportions			
	Predicted		
Actual	0	1	Error
0	66.8	30.5	31.3
1	1.4	1.4	49.3

Overall Accuracy	0.6817
Precision	0.0436
Recall or Sensitivity	0.5060
Specificity TNR	0.6866
Harmonic Mean of Precision and Sensitivity	0.0803
AUC	0.60

Figure 24 Admission (R) Test Curves



(a) ROC (b) P/R

Admission (U, C)

ntree = 500
mtry = 22
Sample Size = 500,500
OOB Estimate of Error = 23.65%

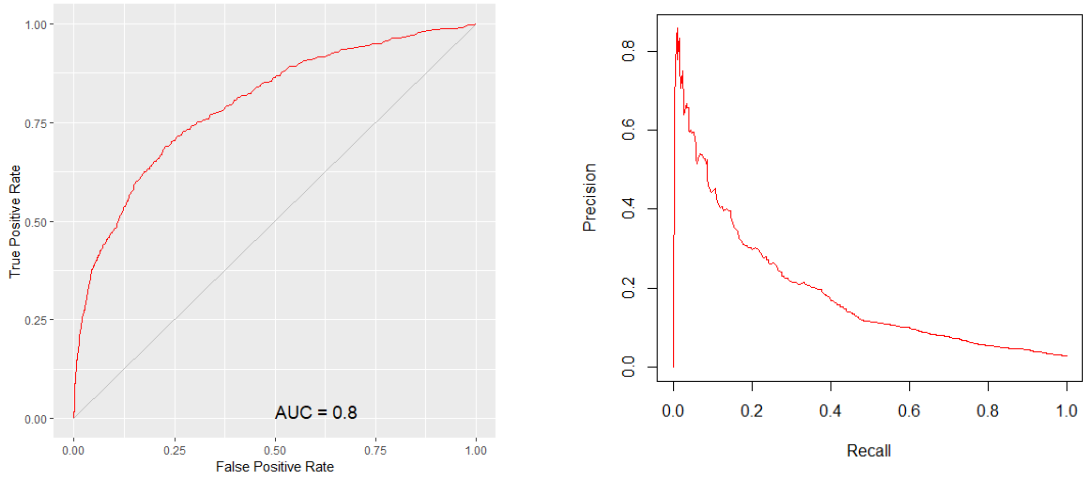
Table 26 Admission (U, C) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	15,688	4,832	23.5
1	179	405	30.7

Proportions			
	Predicted		
Actual	0	1	Error
0	74.3	22.9	23.5
1	0.8	1.9	30.7

Overall Accuracy	0.7626
Precision	0.0773
Recall or Sensitivity	0.6935
Specificity TNR	0.7645
Harmonic Mean of Precision and Sensitivity	0.1392
AUC	0.80

Figure 25 Admission (U, C) Test ROC Curve



(a) ROC (b) P/R

Admission (U, C, P)

ntree = 500
mtry = 24
Sample Size = 500,500
OOB Estimate of Error =

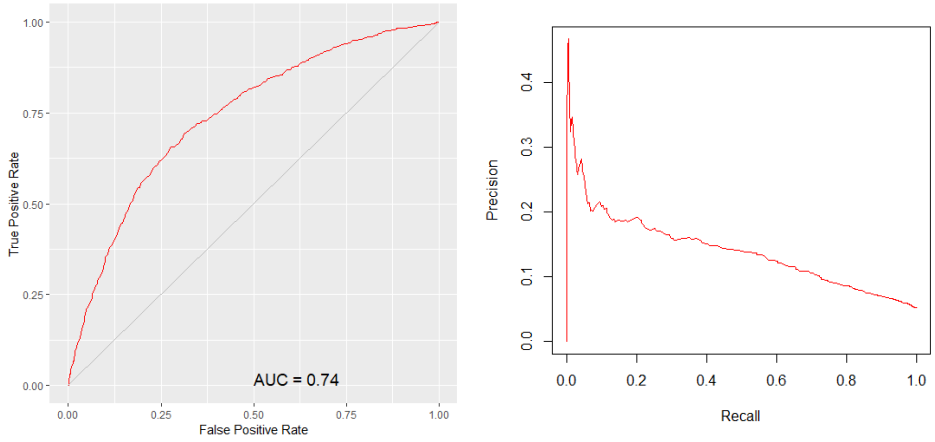
Table 27 Admission (U, C, P) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	5,226	5,294	25.8
1	152	432	26

Proportions			
	Predicted		
Actual	0	1	Error
0	72.1	25.1	25.8
1	0.7	2	26

Overall Accuracy	0.5095
Precision	0.0754
Recall or Sensitivity	0.7397
Specificity TNR	0.4968
Harmonic Mean of Precision and Sensitivity	0.1369
AUC	0.74

Figure 26 Admission (U,C,P) Test Curves



(a) ROC (b) P/R

Admission (U ,C, P, M)

ntree = 500
mtry = 36
Sample Size = 500,500
OOB Estimate of Error = 23.42%

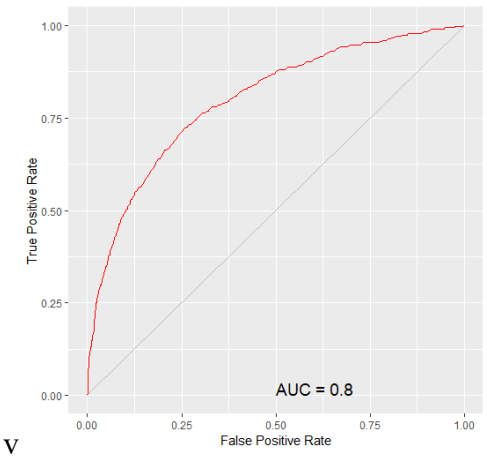
Table 28 Admission (U ,C, P, M) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	15,759	4,761	23.2
1	181	403	31

Proportions			
	Predicted		
Actual	0	1	Error
0	74.7	22.6	23.2
1	0.9	1.9	31

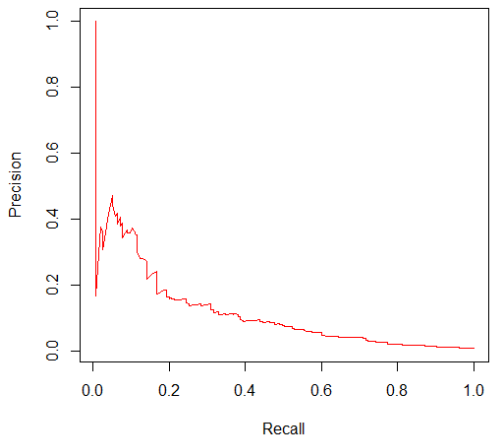
Overall Accuracy	0.7658
Precision	0.0780
Recall or Sensitivity	0.6901
Specificity TNR	0.7680
Harmonic Mean of Precision and Sensitivity	0.1402
AUC	0.80

Figure 27 Admission (U ,C, P, M) Test Curves



V

(a) ROC



(b) P/R

Admission (U, C, P, M, R)

ntree = 500
mtry = 43
Sample Size = 500,500
OOB Estimate of Error = 27.03%

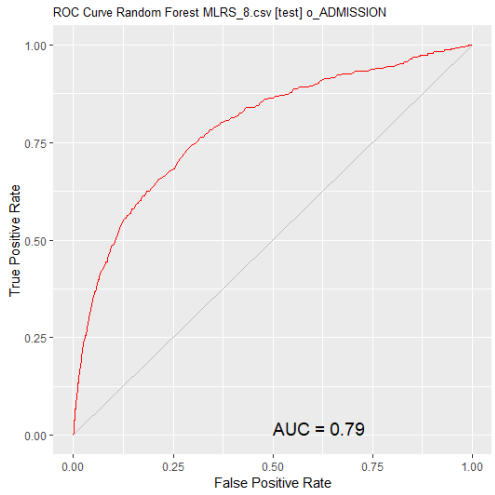
Table 29 Admission (U, C, P, M, R) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	14,906	5,605	27.3
1	164	414	28.4

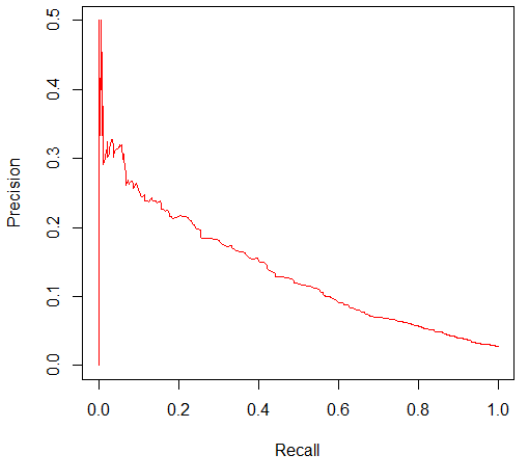
Proportions			
	Predicted		
Actual	0	1	Error
0	70.7	26.6	27.3
1	0.8	2	28.4

Overall Accuracy	0.7264
Precision	0.0688
Recall or Sensitivity	0.7163
Specificity TNR	0.7267
Harmonic Mean of Precision and Sensitivity	0.1255
AUC	0.79

Figure 28 Admission (U, C, P, M, R) Test Curves



(a) ROC



(b) P/R

Readmission (U)

ntree = 500
mtry = 2
Sample Size = 300,300
OOB Estimate of Error = 17.61%

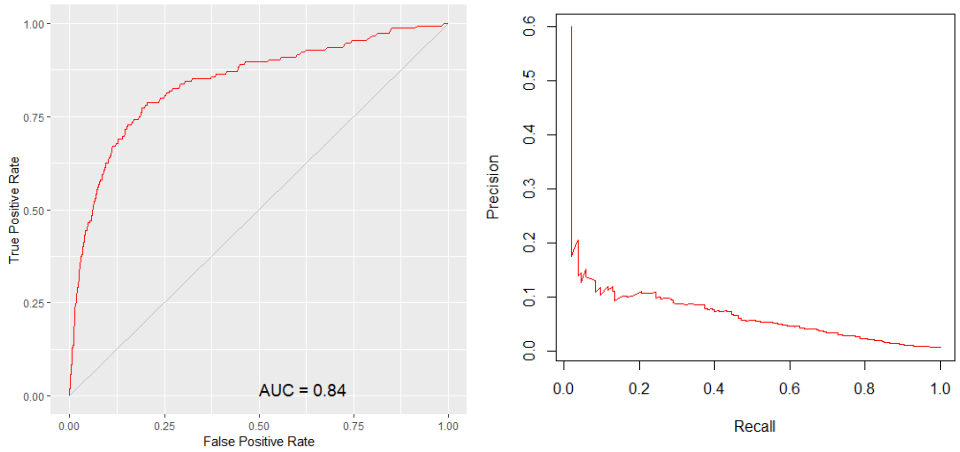
Table 30 Readmission (U) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	17,272	3,677	17.6
1	40	115	25.8

Proportions			
	Predicted		
Actual	0	1	Error
0	81.8	17.4	17.6
1	0.2	0.5	25.8

Overall Accuracy	0.8239
Precision	0.0303
Recall or Sensitivity	0.7419
Specificity TNR	0.8245
Harmonic Mean of Precision and Sensitivity	0.0583
AUC	0.84

Figure 29 Readmission (U) Test Curves



(a) ROC (b) P/R

Readmission (C)

ntree = 500
mtry = 22
Sample Size = 300, 300
OOB Estimate of Error = 16.73%

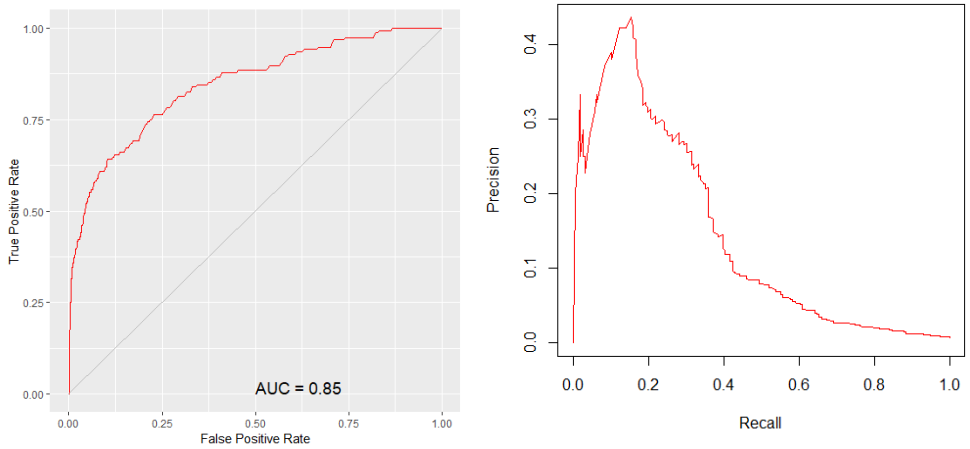
Table 31 Readmission (C) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	17,509	3,439	16.4
1	49	107	31.4

Proportions			
	Predicted		
Actual	0	1	Error
0	83	16.3	16.4
1	0.2	0.5	31.4

Overall Accuracy	0.8347
Precision	0.0302
Recall or Sensitivity	0.6859
Specificity TNR	0.8358
Harmonic Mean of Precision and Sensitivity	0.0578
AUC	0.85

Figure 30 Readmission (C) Test Curves



(a) ROC (b) P/R

Readmission (P)

ntree = 500
mtry = 8
Sample Size = 300, 300
OOB Estimate of Error = 19.98%

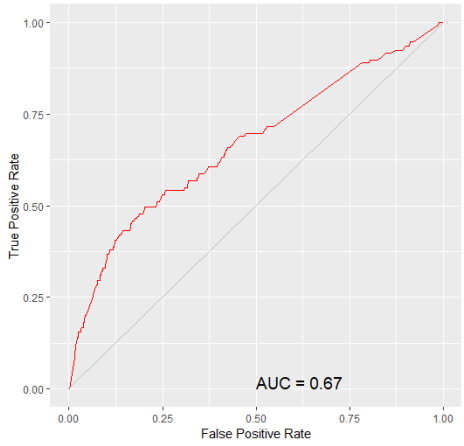
Table 32 Readmission (P) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	16,785	4,164	19.9
1	80	75	51.6

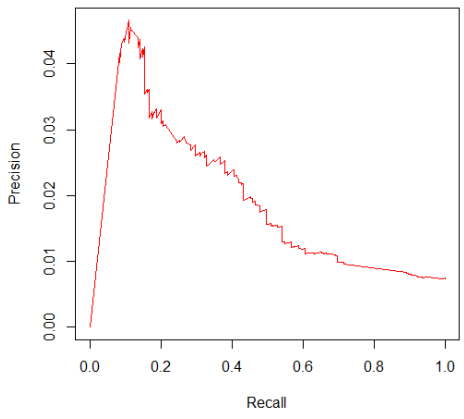
Proportions			
	Predicted		
Actual	0	1	Error
0	79.5	19.7	19.9
1	0.4	0.4	51.6

Overall Accuracy	0.7989
Precision	0.0177
Recall or Sensitivity	0.4839
Specificity TNR	0.8012
Harmonic Mean of Precision and Sensitivity	0.0341
AUC	0.67

Figure 31 Readmission (P) Test ROC Curve



(a) ROC



(b) P/R

Readmission (M)

ntree = 500
mtry = 26
Sample Size = 300, 300
OOB Estimate of Error = 5.99%

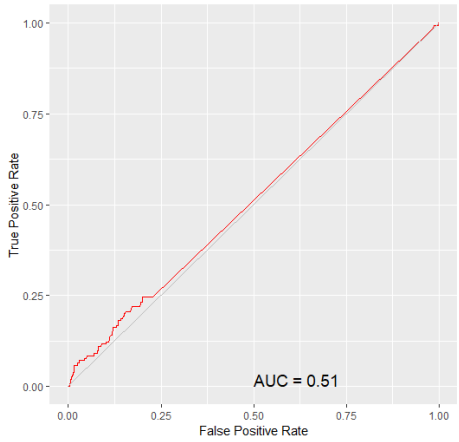
Table 33 Readmission (M) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	19,810	1,139	5.4
1	142	13	91.6

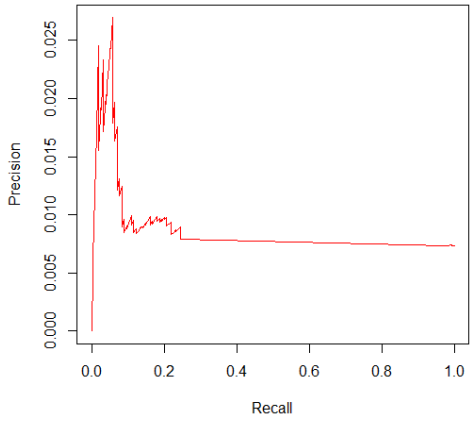
Proportions			
	Predicted		
Actual	0	1	Error
0	91.9	5.4	5.4
1	0.7	0.1	91.6

Overall Accuracy	0.9393
Precision	0.0113
Recall or Sensitivity	0.0839
Specificity TNR	0.9456
Harmonic Mean of Precision and Sensitivity	0.0199
AUC	0.51

Figure 32 Readmission (M) Test Curves



(a) ROC



(b) P/R

Readmission (R)

ntree = 500
mtry = 24
Sample Size = 300, 300
OOB Estimate of Error = 19.62%

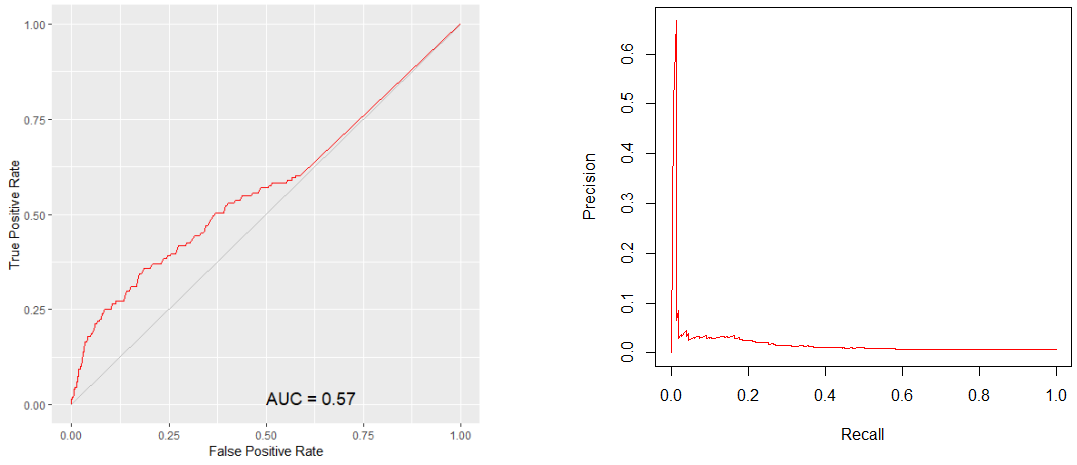
Table 34 Readmission (R) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	16,845	4,093	19.5
1	97	54	64.2

Proportions			
	Predicted		
Actual	0	1	Error
0	79.9	19.4	19.5
1	0.5	0.3	64.2

Overall Accuracy	0.8013
Precision	0.0130
Recall or Sensitivity	0.3576
Specificity TNR	0.8045
Harmonic Mean of Precision and Sensitivity	0.0251
AUC	0.57

Figure 33 Readmission (R) Test ROC Curve



(a) ROC (b) P/R

Readmission (U, C)

ntree = 500
mtry = 22
Sample Size = 300, 300
OOB Estimate of Error = 16.67%

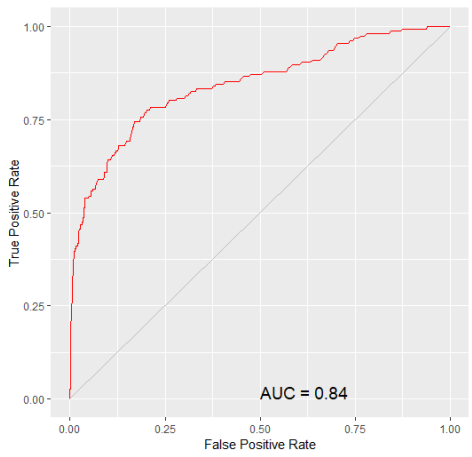
Table 35 Readmission (U,C) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	17,477	3,471	16.6
1	42	114	26.9

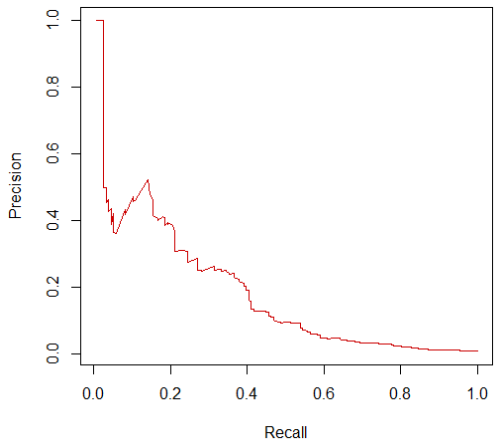
Proportions			
	Predicted		
Actual	0	1	Error
0	82.8	16.4	16.6
1	0.2	0.5	26.9

Overall Accuracy	0.8335
Precision	0.0318
Recall or Sensitivity	0.7308
Specificity TNR	0.8343
Harmonic Mean of Precision and Sensitivity	0.0609
AUC	0.84

Figure 34 Readmission (U,C) Test Curves



(a) ROC



(b) P/R

Readmission (U, C, P)

ntree = 500
mtry = 24
Sample Size = 300,300
OOB Estimate of Error = 15.59%

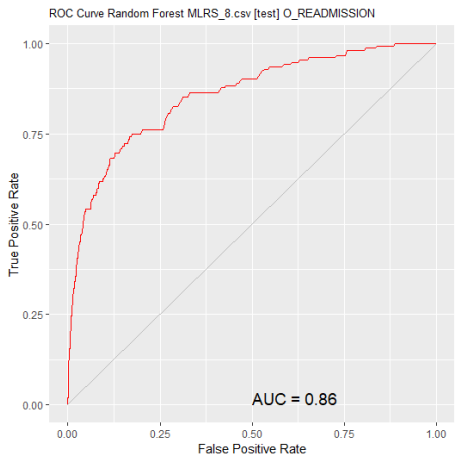
Table 36 Readmission (U, C, P) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	17,276	3,673	17.5
1	41	114	26.5

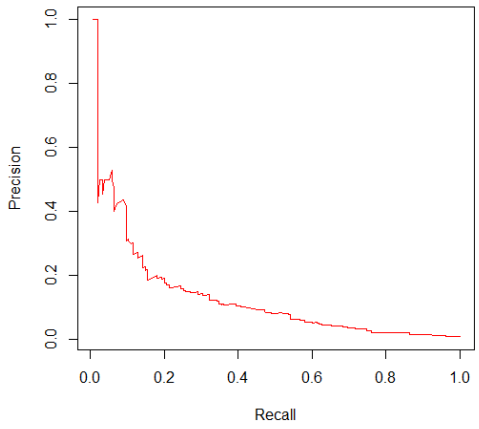
Proportions			
	Predicted		
Actual	0	1	Error
0	81.9	17.4	17.5
1	0.2	0.5	26.5

Overall Accuracy	0.8240
Precision	0.0301
Recall or Sensitivity	0.7355
Specificity TNR	0.8247
Harmonic Mean of Precision and Sensitivity	0.0578
AUC	0.85

Figure 35 Readmission (U, C, P) Test Curves



(a) ROC



(b) P/R

Readmission (U, C, P, M)

ntree = 500
mtry = 36
Sample Size = 500
OOB Estimate of Error = 15.54%

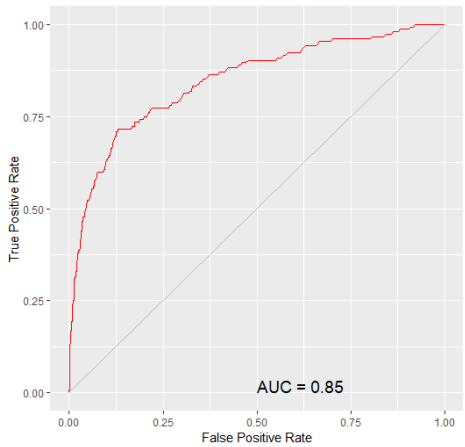
Table 37 Readmission (U, C, P, M) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	17,791	3,158	15.1
1	44	111	28.4

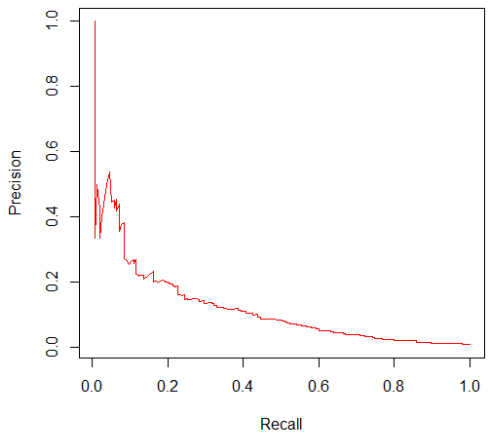
Proportions			
	Predicted		
Actual	0	1	Error
0	84.3	15	15.1
1	0.2	0.5	28.4

Overall Accuracy	0.8483
Precision	0.0340
Recall or Sensitivity	0.7161
Specificity TNR	0.8493
Harmonic Mean of Precision and Sensitivity	0.0648
AUC	0.85

Figure 36 Readmission (U, C, P, M) Test ROC Curve



(a) ROC



(b) P/R

Readmission (U, C, P, M, R)

ntree = 500
mtry = 43
Sample Size = 300,300
OOB Estimate of Error = 16.94%

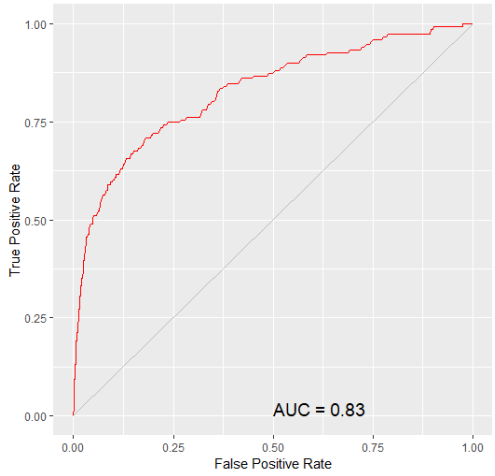
Table 38 Readmission (U, C, P, M, R) Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	17,403	3,535	16.9
1	48	103	31.8

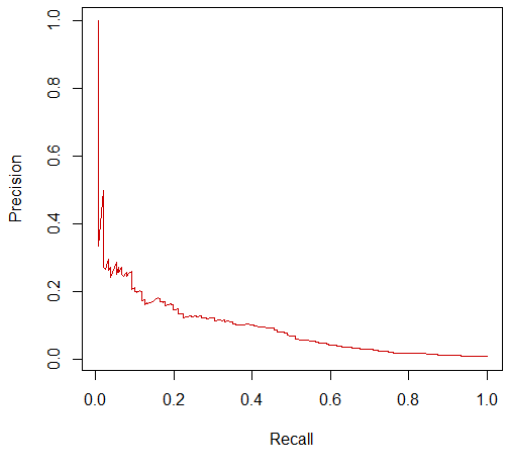
Proportions			
	Predicted		
Actual	0	1	Error
0	82.5	16.8	16.9
1	0.2	0.5	31.8

Overall Accuracy	0.8301
Precision	0.0283
Recall or Sensitivity	0.6821
Specificity TNR	0.8312
Harmonic Mean of Precision and Sensitivity	0.0544
AUC	0.83

Figure 37 Readmission (U, C, P, M, R) Test ROC Curve



(a) ROC



(b) P/R

Multi-ED (U)

ntree = 500
mtry = 2
Sample Size = 500,500
OOB Estimate of Error = 33.33%

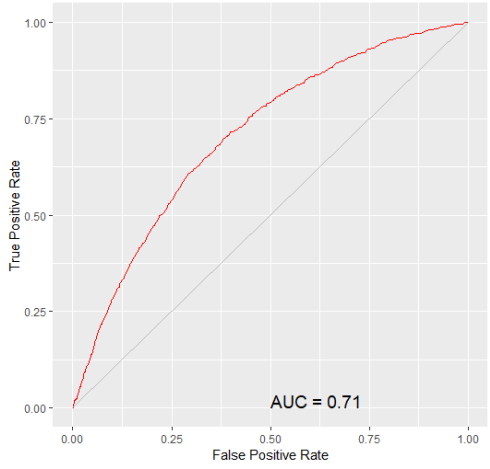
Table 39 Multi-ED Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,411	6,597	33
1	391	705	35.7

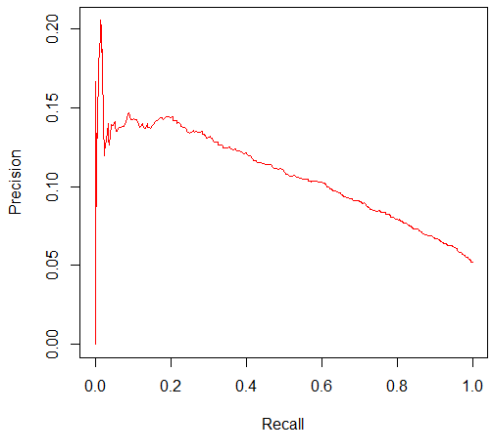
Proportions			
	Predicted		
Actual	0	1	Error
0	69.5	31.3	33
1	1.9	3.3	35.7

Overall Accuracy	0.6689
Precision	0.0965
Recall or Sensitivity	0.6432
Specificity TNR	0.6703
Harmonic Mean of Precision and Sensitivity	0.1679
AUC	0.71

Figure 38 Multi-ED (U) Test Curves



(a) ROC



(b) P/R

Multi-ED (C)

ntree = 500
mtry = 22
Sample Size = 500, 500
OOB Estimate of Error = 31.41%

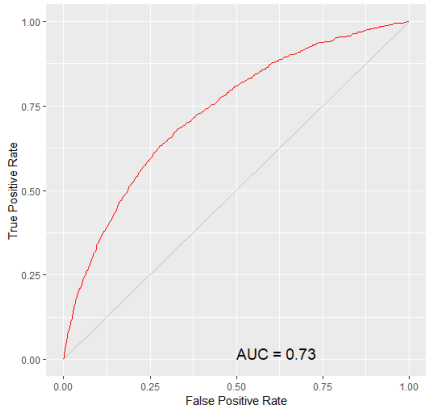
Table 40 Multi-ED Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,749	6,259	31.3
1	375	721	34.2

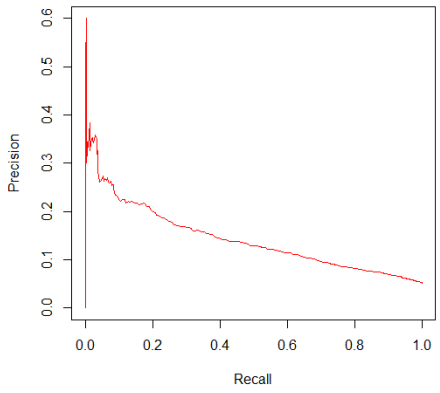
Proportions			
	Predicted		
Actual	0	1	Error
0	65.1	29.7	31.3
1	1.8	3.4	34.2

Overall Accuracy	0.6857
Precision	0.1033
Recall or Sensitivity	0.6578
Specificity TNR	0.6872
Harmonic Mean of Precision and Sensitivity	0.1786
AUC	0.73

Figure 39 Multi-ED (C) Test Curves



(a) ROC



(b) P/R

Multi-ED (P)

ntree = 500
mtry = 8
Sample Size = 500, 500
OOB Estimate of Error = 28.3%

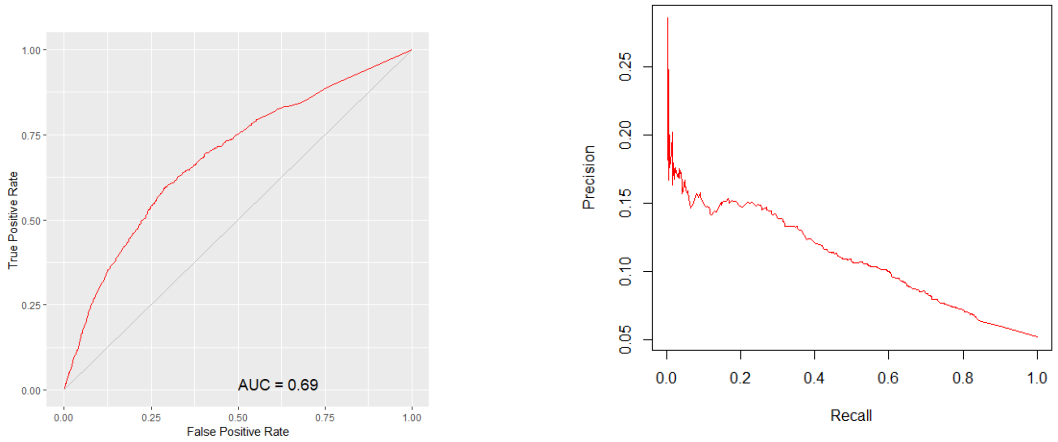
Table 41 Multi-ED (P) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	14,558	5,450	27.2
1	471	625	43

Proportions			
	Predicted		
Actual	0	1	Error
0	69	25.8	27.2
1	2.2	3	43

Overall Accuracy	0.7194
Precision	0.1029
Recall or Sensitivity	0.5703
Specificity TNR	0.7276
Harmonic Mean of Precision and Sensitivity	0.1743
AUC	0.69

Figure 40 Multi-ED Test Curves



(a) ROC (b) P/R

Multi-ED (M)

ntree = 500
mtry = 26
Sample Size = 500, 500
OOB Estimate of Error = 12.58%

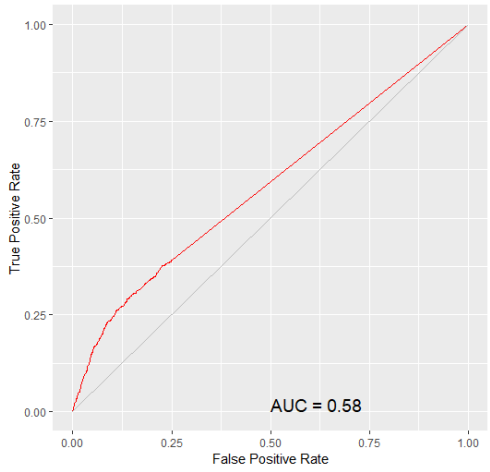
Table 42 Multi-ED Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	18,143	1,865	9.3
1	839	257	76.6

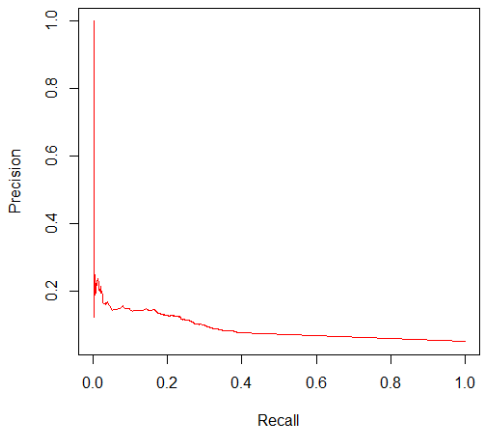
Proportions			
	Predicted		
Actual	0	1	Error
0	86	8.8	9.3
1	4	1.2	76.6

Overall Accuracy	0.8719
Precision	0.1211
Recall or Sensitivity	0.2345
Specificity TNR	0.9068
Harmonic Mean of Precision and Sensitivity	0.1597
AUC	0.58

Figure 41 Multi-ED (M) Test Curves



(a) ROC



(b) P/R

Multi-ED (R)

ntree = 500
mtry = 24
Sample Size = 500, 500
OOB Estimate of Error = 26.54%

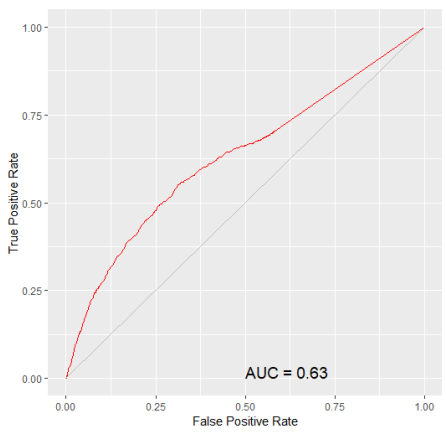
Table 43 Multi-ED (R) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	15,010	4,987	24.9
1	570	522	52.2

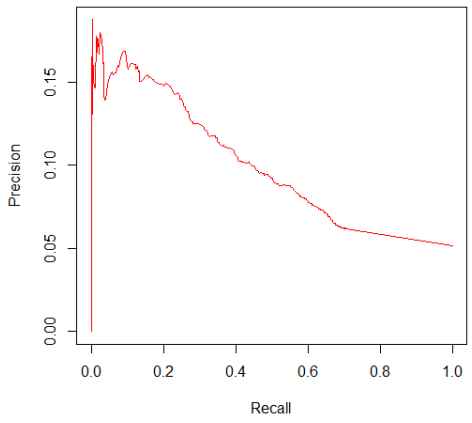
Proportions			
	Predicted		
Actual	0	1	Error
0	71.2	23.6	24.9
1	2.71	2.5	52.2

Overall Accuracy	0.7365
Precision	0.0948
Recall or Sensitivity	0.4780
Specificity TNR	0.7506
Harmonic Mean of Precision and Sensitivity	0.1582
AUC	0.63

Figure 42 Multi-ED (R) Test ROC Curve



(a) ROC



(b) P/R

Multi-ED (U, C)

ntree = 500
mtry = 22
Sample Size = 500, 500
OOB Estimate of Error = 31.86

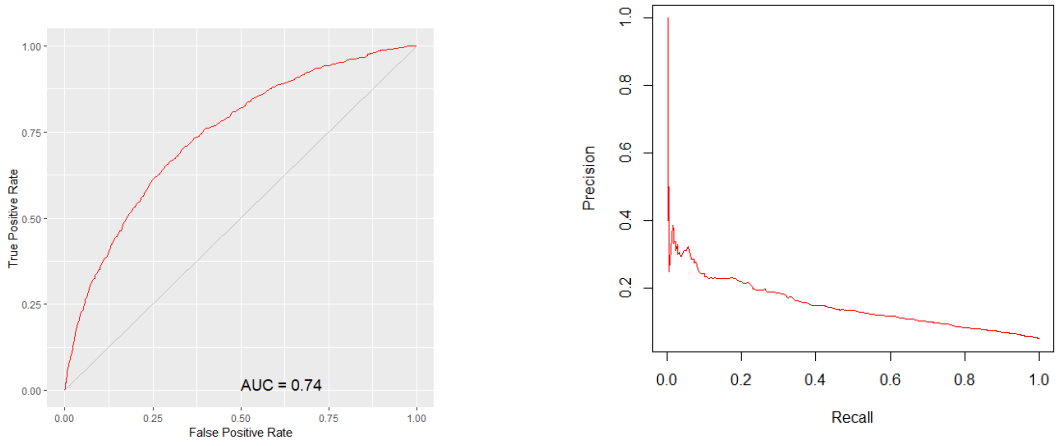
Table 44 Multi-ED (U,C) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,660	6,354	31.7
1	348	742	31.9

Proportions			
	Predicted		
Actual	0	1	Error
0	64.7	30.1	31.7
1	1.6	3.5	31.9

Overall Accuracy	0.6824
Precision	0.1046
Recall or Sensitivity	0.6807
Specificity TNR	0.6825
Harmonic Mean of Precision and Sensitivity	0.1813
AUC	0.74

Figure 43 Multi-ED (U,C) Test ROC Curve



(a) ROC (b) P/R

Multi-ED (U,C,P)

ntree = 500
mtry = 43
Sample Size = 300,300
OOB Estimate of Error = 33.63%

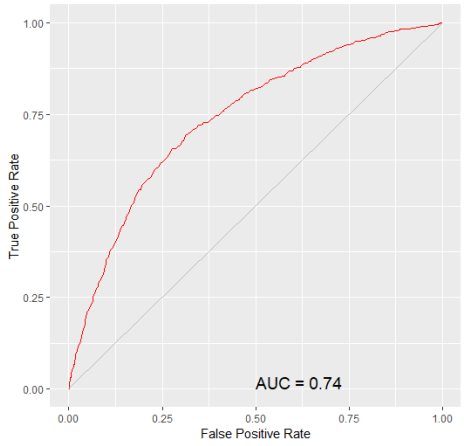
Table 45 Multi-ED (U,C,P) Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	13,679	6,329	31.6
1	342	754	31.2

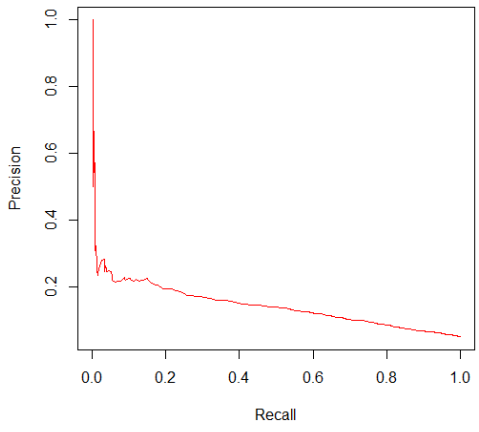
Proportions			
	Predicted		
Actual	0	1	Error
0	64.8	30	31.6
1	1.6	3.6	31.2

Overall Accuracy	0.6839
Precision	0.1065
Recall or Sensitivity	0.6880
Specificity TNR	0.6837
Harmonic Mean of Precision and Sensitivity	0.1844
AUC	0.74

Figure 44 Multi-ED (U,C,P) Test Curves



(a) ROC



(b) P/R

Multi-ED (U, C, P, M)

ntree = 500
mtry = 36
Sample Size = 500,500
OOB Estimate of Error = 29.99%

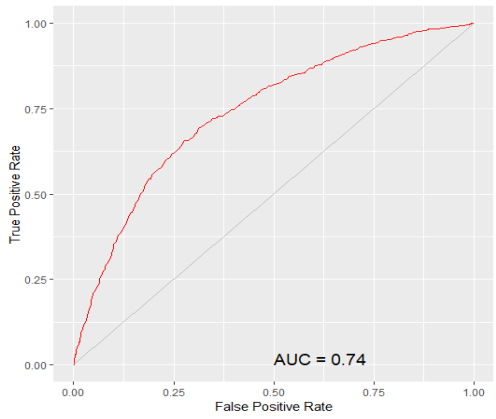
Table 46 Multi-ED (U, C, P, M) Test Error Matrix

Counts			
	Predicted		
Actual	0	1	Error
0	14,066	5,942	29.7
1	313	723	34

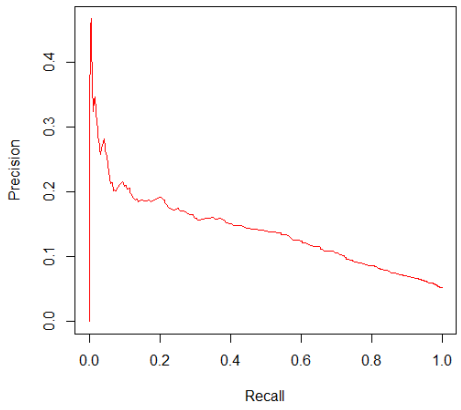
Proportions			
	Predicted		
Actual	0	1	Error
0	66.7	28.2	29.7
1	1.8	3.4	34

Overall Accuracy	0.7028
Precision	0.1085
Recall or Sensitivity	0.6979
Specificity TNR	0.7030
Harmonic Mean of Precision and Sensitivity	0.1878
AUC	0.74

Figure 45 Multi-ED (U, C, P, M) Test Curves



(a) ROC



(b) P/R

Multi-ED (U, C, P, M, R)

ntree = 500
mtry = 43
Sample Size = 500,500
OOB Estimate of Error = 31.1%

Table 47 Multi-ED (U, C, P, M, R) Test Error Matrix

Counts

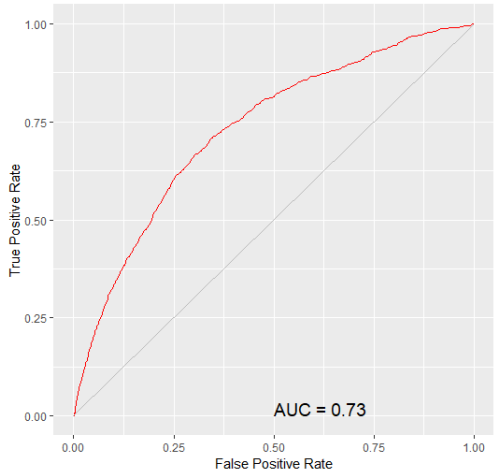
	Predicted		
Actual	0	1	Error
0	13,854	6,143	30.7
1	363	729	33.2

Proportions

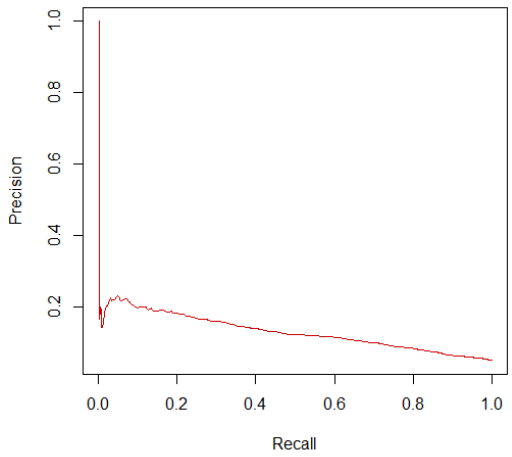
	Predicted		
Actual	0	1	Error
0	65.7	29.1	30.7
1	1.7	3.5	33.2

Overall Accuracy	0.6915
Precision	0.1061
Recall or Sensitivity	0.6676
Specificity TNR	0.6928
Harmonic Mean of Precision and Sensitivity	0.1831
AUC	0.73

Figure 46 Multi-ED (U, C, P, M, R) Test ROC Curve



(a) ROC



(b) P/R

Chapter V

DISCUSSION

The results show that the data and Random Forest models can be used to predict the outcomes of interest better than chance to a varying degree. The UCPM model was the best performing with highest AUC scores for three of the four outcomes. The UCP model slightly outperformed with a AUC of 0.86 for the readmission compared to an AUC of 0.85 for the UCPM model. The strongest individual data type for predicting high cost and admission was the U data type with an AUC of 0.80 for both outcomes. The strongest individual data type for predicting both readmission and multiple emergency visits was the C data type with 0.85 and 0.73 respectively. The combined models UCP had the highest AUC score of 0.86 for the readmission outcome. Among these models the highest F_1 score was the UCPM High Cost model with a F_1 score of 0.3291. When predicting hospital admission, U alone produced an AUC of 0.80, the same as UC and UCPM with the UCPM producing a slightly higher F_1 score of 0.1402. See Figure 9 (a) for a comparison of the AUC for each model and outcome and Figure 9 (e) for F_1 scores.

The weakest individual data types were R and M. Introducing the R data type to the model with other data types resulted in a moderate decrease in the ROC AUC scores for all four outcomes. The M readmission model had an overall accuracy of 0.9393 but this was a result of the weak model (ROC 0.51) correctly predicting the negative outcome 19,810 times, an error rate of 5.4 and correctly predicting the positive outcome 13 times an error rate of 91.6 (see Table 33). The overall error was 6% but the average class error was 48.5%.

All of the F_1 scores were relatively low. Recall was higher than precision in all models and highest recall ranged between 0.7099 – 0.7514 across the four outcomes. It is important to note, while precision was lower than recall, the business use case for care management would generally benefit from a higher recall than a model with higher precision and lower recall.

Limitations of the Data

The data for conditions, procedures, and medications were designated as binary features for this study. Alternative approaches may provide richer information for these data types. Some frequency or magnitude could provide better predictive power as compared to a positive or negative indicator. The results portion of the dataset used the result value from the maximum date available in the baseline timeframe. This may exclude important result data from multiple values recorded during the baseline timeframe. It may be beneficial to incorporate

A bias has been identified in discharge coding where a patient's severity of illness inversely affects the coding of some common conditions that on their own are not life

threatening but cumulatively increase risk for adverse outcomes including incurring high cost[38,78].

Potential Sources of Error in Data

The data is collected from many systems in many formats. This is a strength in that the data is broad however, it could become a weakness if the accuracy of the longitudinal record is diminished by significant noise in the data. Every time a data point is mapped to a concept there is a chance that the meaning of the data can be distorted or specificity altered. Patient matching is another potential source of error. As contributing data system use different and separate patient identifiers. The patient matching process is both deterministic and probabilistic however patient matching errors can occur resulting in inaccurate relationships between model features and outcomes.

In real-world applications patients often have periods of non-continuous coverage or change in attribution state. The study only included patients with 24 months of continuous coverage in a contract. Roughly 40% of the total ACO population met the continuous enrollment inclusion criteria. While this impacted study inclusion, the resulting models could perform similarly in predicting future outcomes with sufficient baseline data. In real-world application, insufficient data will continue to be a challenge when a patient is new to treatment environment and contributing data sources.

Considerations for Future Research

For future research it may be beneficial to enhance the feature sets in a number of ways. Capturing the frequency of procedures and the dosage and frequency of

medications may strengthen the predictive power of these data types. It may also be beneficial to develop a result feature set that captures multiple results and incorporates result reference ranges.

The mean decrease in Gini Index measurement used to rank variable importance as described by Breiman and deployed in the RandomForest R package is biased towards categorical features with many categories[62,73,74]. In the future, variable importance could be measured using Conditional Inference trees[73,74]. This could potentially reduce the number of features used in the models.

Additional outcomes of interest could be explored with the input features developed for this study. Pre-existing diagnosis groupers such as the Charlson co-morbidity index, Elixhauser co-morbidity index, or Hierarchical Condition Categories model could be incorporated into the input features as a composite feature.

A subset of the population had positive outcomes in all four of the targets in this study. Future research may look to predict which patients will incur all four of these outcomes.

Chapter VI

SUMMARY AND CONCLUSION

This research demonstrates that disparate data sources and EDW ontology mappings can be used as features along with random forest models to predict population health focused outcomes. Other risk adjustment tools use claims, and diagnosis data to predict cost or adjust payments. This research is novel because it uses many data sources and population health EDW cross ontology mappings as features for the random forest models. The results show that similar to prior risk adjusters, demographics and diagnoses were strong predictors of high cost. Results and medications alone were the least predictive of the target outcomes. Future research may leverage similar data assets from across a care continuum. The framework used in this research has the potential to expand and scale to include any number of additional data types and outcomes. As reimbursement models continue to shift, clinically integrated networks and integrated delivery systems must find ways to reduce waste and improve outcomes and cost. Continuing to find new and improved methods for identifying patients who will have

concentrated cost and services is an important task for those who perform care management services.

References

1. Davis K, Stremikis K, Squires D, Schoen C, Commonwealth Fund. Mirror, mirror on the wall : how the performance of the U.S. health care system compares internationally : 2014 update. New York, NY: Commonwealth Fund,; 2014:1 online resource (PDF file (31 pages)).

2. Capretta JC. The role of medicare fee-for-service in inefficient health care delivery: American Enterprise Institute for Public Policy Research; 2013.

3. Kelley R. Where can $700 billion in waste be cut annually from the US healthcare system. Ann Arbor, MI: Thomson Reuters 2009;24.

4. Moore HL. Cours d'Économie Politique. By VILFREDO PARETO, Professeur à l'Université de Lausanne. Vol. I. Pp. 430. I896. Vol. II. Pp. 426. I897. Lausanne: F. Rouge. Annals of the American Academy of Political & Social Science 1897;9:128.

5. Kanjirathinkal M. Pareto Analyzes the Distribution of Wealth. Salem Press; 2013.

6. Cohen SB. The concentration of health care expenditures in the U.S. and predictions of future spending. Journal of Economic & Social Measurement 2016;41:167-89.

7. Arnold DR. Risk Stratifying a Population for the Comprehensive Primary Care Plus (CPC+) Program. Rutgers University; 2017.

8. Anderson GF. Chronic care: making the case for ongoing care: Robert Wood Johnson Foundation; 2010.

9. Anderson GF. Medicare and chronic conditions. Mass Medical Soc; 2005.

10. Comprehensive Primary Care Plus. 2017. (Accessed 11/12/2017, 2017, at https://innovation.cms.gov/initiatives/comprehensive-primary-care-plus.)

11. Lee J, Anderson T. High-cost Medicare Beneficiaries. 2005: Congress of the United States, Congressional Budget Office.

12. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Affairs 2008;27:759-69.

13. McClellan MB, Leavitt MO. Competencies and tools to shift payments from volume to value. Jama 2016;316:1655-6.

14. Hileman G, Steele S. Accuracy of Claims-Based Risk Scoring Models: Society of Actuaries; 2016.

15. Bodenheimer T, Berry-Millett R. Care management of patients with complex health care needs. Policy 2009;1:6.

16. Johnson TL, Brewer D, Estacio R, et al. Augmenting Predictive Modeling Tools with Clinical Insights for Care Coordination Program Design and Implementation. eGEMs 2015;3.

17. Merriam-Webster. Merriam-Webster's collegiate dictionary: Merriam-Webster; 2004.

18. Iezzoni LI, NetLibrary Inc. Risk adjustment for measuring health care outcomes. 3rd ed. Chicago, Ill.: Health Administration Press; 2003:xx, 508 p.

19. Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes: AUPHA; 2013.

20. Hong C, Hwang A, Ferris T. Finding a Match: How Successful Complex Care Programs Identify Patients. California Health Care Foundation March 2015.

21. Hong CS, Siegel AL, Ferris TG. Caring for high-need, high-cost patients: what makes for a successful care management program. Issue Brief (Commonw Fund) 2014;19:9.

22. Haas LR, Takahashi PY, Shah ND, et al. Risk-stratification methods for identifying patients for care coordination. American Journal of Managed Care 2013;19:725-32.

23. Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes. 3rd ed. Chicago, Ill.: Health Administration Press; 2003:xx, 508 p.

24. Bindman AB, Grumbach K, Osmond D, et al. Preventable hospitalizations and access to health care. Jama 1995;274:305-11.

25. Drewnowski A, Rehm CD, Solet D. Disparities in obesity rates: analysis by ZIP code area. Social science & medicine 2007;65:2458-63.

26. Acevedo-Garcia D. Zip code-level risk factors for tuberculosis: neighborhood environment and residential segregation in New Jersey, 1985-1992. American Journal of Public Health 2001;91:734.

27. Gould JM. Quality of life in American neighborhoods. Levels of affluence, toxic waste, and cancer mortality in residential zip code areas. 1986.

28. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. American Journal of Public Health 2005;95:312-23.

29. Pope GC, Kautter J, Ellis RP, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financing Review 2004;25:119-41.

30. Charlson M, Charlson RE, Briggs W, Hollenberg J. Can disease management target patients most likely to generate high costs? The impact of comorbidity. Journal of general internal medicine 2007;22:464-9.

31. Charlson M, Wells MT, Ullman R, King F, Shmukler C. The Charlson comorbidity index can be used prospectively to identify patients who will incur high future costs. PloS one 2014;9:e112479.

32. Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. J Clin Epidemiol 2008;61:1234-40.

33. Li B, Evans D, Faris P, Dean S, Quan H. Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. BMC Health Services Research 2008;8.

34. Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chron Dis 1987; 40: 373–383. External Resources Pubmed/Medline (NLM) CrossRef (DOI) Chemical Abstracts Service (CAS) 1985.

35. Charlson M, Wells MT, Ullman R, King F, Shmukler C. The Charlson comorbidity index can be used prospectively to identify patients who will incur high future costs. PLoS ONE 2014;9.

36. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45:613-9.

37. Hude Quan a, Vijaya Sundararajan a, Patricia Halfon a, et al. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. Med Care 2005:1130.

38. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998;36:8-27.

39. Robert B. Fetter a, Youngsoo Shin a, Jean L. Freeman a, Richard F. Averill a, John D. Thompson a. Case Mix Definition by Diagnosis-Related Groups. Med Care 1980:i.

40. Hsiao WC, Sapolsky HM, Dunn DL, Weiner SL. Lessons of the New Jersey DRG payment system. Health affairs (Project Hope) 1986;5:32-45.

41. Averill RF, Muldoon JH, Vertrees JC, et al. The evolution of casemix measurement using diagnosis related groups (DRGs). Wallingford: 3M Health Information Systems 1998.

42. Medicare program; changes to the hospital inpatient prospective payment systems and fiscal year 2008 rates. Fed Regist 2007;72:47129-8175.

43. Fries BE, Cooney Jr LM. Resource utilization groups: A patient classification system for long-term care. Med Care 1985:110-22.

44. Fries BE, Schneider DP, Foley WJ, Gavazzi M, Burke R, Cornelius E. Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III). Med Care 1994:668-85.

45. Nursing home resident assessment : resource utilization groups. Washington, D.C.: Department of Health and Human Services, Office of Inspector General; 2001.

46. Averill RF, Goldfield NI, Wynn ME, et al. Design of a prospective payment patient classification system for ambulatory care. Health care financing review 1993;15:71.

47. Goldfield N, Averill R, Eisenhandler J, Grant T. Ambulatory patient groups, version 3.0ĝ€"A classification system for payment of ambulatory visits. Journal of Ambulatory Care Management 2008;31:2-16.

48. Medicare Cf, Medicaid Services H. Medicare program: changes to the hospital inpatient prospective payment systems and fiscal year 2009 rates; payments for graduate medical education in certain emergency situations; changes to disclosure of physician ownership in hospitals and physician self-referral rules; updates to the long-term care prospective payment system; updates to certain IPPS-excluded hospitals; and collection of information regarding financial relationships between hospitals. Final rules. Federal Register 2008;73:48433.

49. Schlenker RE, Powell MC, Goodrich GK. Initial home health outcomes under prospective payment. Health Services Research 2005;40:177-93.

50. Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2017. Final rule. Federal register 2016;81:52055-141.

51. U.S. Centers for Medicare & Medicaid Services. National Health Expenditures by type of service and source of funds, CY 1960-2015. In: Services USDoHaH, ed. 7500 Security Boulevard, Baltimore, MD 212442017.

52. Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A. Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financing Review 1989;10:17.

53. Yi R, Shreve JL, Bluhm WF. Risk adjustment and its Applications in global payments to providers. Risk 2011;2.

54. Eisenhandler J, Averil R, Vertrees J. A comparison of the explanatory power of two approaches to the prediction of post acute care resources use: Technical Report. 2011. 3M Health Information Systems. This is a CMS-commissioned report comparing the predictive value of CRG vs. the Diagnostic Cost Group Hierarchical Clinical Conditions (HCC) risk adjuster; 2015.

55. Systems MHI. 3M Ckinical Risk groups: Measuring Risk, Managing Care. Salt Lake City, UT 3M; 2016.

56. Hughes JS, Averill RF, Eisenhandler J, et al. Clinical Risk Groups (CRGs) a classification system for risk-adjusted capitation-based payment and health care management. Med Care 2004;42:81-90.

57. Draaghtel KL, Diane. Risk Adjustment and the Power of Four. Milliman; 2012.

58. Evans MA, Pope GC, Kautter J, et al. Evaluation of the CMS-HCC Risk Adjustment Model. CfMM Services, Editor 2011.

59. Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. Soviet physics doklady; 1966. p. 707-10.

60. Corporation C. Master Person Management. Kansas City MO: Cerner Corporation; 2017.

61. Cost H, Project U. HCUP quality control procedures. Rockville, MD: Agency for Healthcare Research and Quality 2016.

62. Breiman L. Random forests. Machine learning 2001;45:5-32.

63. Williams G. Data mining with Rattle and R: The art of excavating data for knowledge discovery: Springer Science & Business Media; 2011.

64. Efron B, Tibshirani RJ. An introduction to the bootstrap: CRC press; 1994.

65. Genuer R, Poggi J-M, Tuleau C. Random Forests: some methodological insights. 2008.

66. Díaz-Uriarte R, De Andres SAJBb. Gene selection and classification of microarray data using random forest. 2006;7:3.

67. Liaw A, Wiener MJRn. Classification and regression by randomForest. 2002;2:18-22.

68. De'ath G, Fabricius KEJE. Classification and regression trees: a powerful yet simple technique for ecological data analysis. 2000;81:3178-92.

69. Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Medical Informatics and Decision Making 2011;11:51-.

70. Volscho TW. Gini Coefficient. 2008:407-8.

71. Biau G. Analysis of a random forests model. Journal of Machine Learning Research 2012;13:1063-95.

72. Dietterich TGJML. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. 2000;40:139-57.

73. Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis AJBb. Conditional variable importance for random forests. 2008;9:307.

74. Strobl C, Boulesteix A-L, Zeileis A, Hothorn TJBb. Bias in random forest variable importance measures: Illustrations, sources and a solution. 2007;8:25.

75. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006;27:861-74.

76. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS one 2015;10:e0118432-e.

77. Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J Eval Clin Pract 2006;12:132-9.

78. Hughes JS, Iezzoni LI, Daley J, Greenberg L. How severity measures rate hospitalized patients. Journal of General Internal Medicine 1996;11:303-11.
The RandomForest package in R defaults the number of