Identifying Key Factors Associated with High Risk Asthma Patients to Reduce the Cost of Health Resources Utilization

Amani Ahmad
Louisiana State University and Agricultural and Mechanical College, amani.ahmad91@gmail.com

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Categorical Data Analysis Commons, Health Services Research Commons, Industrial Engineering Commons, Multivariate Analysis Commons, and the Statistical Methodology Commons

Recommended Citation
Ahmad, Amani, "Identifying Key Factors Associated with High Risk Asthma Patients to Reduce the Cost of Health Resources Utilization" (2018). LSU Master's Theses. 4803.
https://digitalcommons.lsu.edu/gradschool_theses/4803
IDENTIFYING KEY FACTORS ASSOCIATED WITH HIGH RISK ASTHMA PATIENTS TO REDUCE THE COST OF HEALTH RESOURCES UTILIZATION

A Thesis

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Masters of Science in Industrial Engineering

in
The Department of Mechanical and Industrial Engineering

by
Amani Ahmad
B.S., University of Jordan, 2014
December 2018
## Table of Contents

Abstract .......................................................................................................................... iv

Chapter 1. Introduction .................................................................................................. 1
1.1 Study motivation .................................................................................................... 3
1.2 Research objectives ............................................................................................... 3

Chapter 2. Literature Review ....................................................................................... 4
2.1 Asthma definition and statistics .......................................................................... 4
2.2 Health resources utilization statistics and cost ................................................... 5
2.3 Factors associated with higher health resources utilization ............................... 6
2.4 Intervention programs ......................................................................................... 7
2.5 High-risk patients definition from previous studies ......................................... 10
2.6 Previous studies data sources ........................................................................... 11
2.7 Research gap ....................................................................................................... 12

Chapter 3. Research Methodology ............................................................................. 14
3.1 Study objective .................................................................................................... 14
3.2 Study setting ....................................................................................................... 14
3.3 Ethical consideration ........................................................................................... 14
3.4 Study population inclusion and exclusion criteria ............................................. 14
3.5 Data collection .................................................................................................... 15
3.6 Study population descriptive statistics ............................................................. 17

Chapter 4. Results ...................................................................................................... 20
4.1 Difference in cost between patients who are case managed vs non-case managed .......................................................................................................................... 20
4.2 Exploring the effect of case management on cost over time ............................... 24
4.3 Forward variable selection procedure for total cost .......................................... 25
4.4 Identifying factors associated with frequent use of health services ............... 29
4.5 Neural networks (NN) vs linear regression in prediction of total cost ............ 32

Chapter 5. Discussion and Limitations ...................................................................... 35

Chapter 6. Conclusion ................................................................................................. 37
Abstract

Asthma is associated with frequent use of primary health services and places a burden on the United States economy. Identifying key factors associated with increased cost of asthma is an essential step to improve practices of asthma management.

The aim of this study was to identify factors associated with over utilization of primary health services and increased cost via claims data and to explore the effectiveness of case management program in reducing overall asthma related cost.

Claims data analysis for Medicaid insured asthma patients in Louisiana was conducted. Asthma patients were identified using their ICD-9 and ICD-10 codes, forward variable selection was used to identify significant factors in the regression model with total cost as the dependent variable, multivariate regression was used to identify patients’ factors associated with frequent utilization of primary health services, and finally, a T-test was used to compare the difference in cost over time for case managed and non-case managed patients.

Cost of four claims categories was significant to the total cost variable: primary physician visits, pharmacy prescriptions, emergency room visits and urgent care clinics visits. Median income and enrollment in case management were significant in predicting number of emergency room visits. Patients who had higher income were more likely to utilize urgent care clinics. As a side finding, this study built a prediction model for total cost, the linear regression model accuracy was compared to neural networks and the proposed threshold point in which neural network outperforms the regression model is around 6,000 data points.

Patients with a history of utilization of certain health services are more likely to need case management for better health outcomes and controlled cost. Future work is to perform analysis on a larger scale and include more patients related factors to identify a more holistic definition of high-risk patients.
Chapter 1. Introduction

Healthcare is one of the largest industries in the world. The United States has the highest healthcare expenses as compared to other countries and the US Health System ranks last among eleven developed countries on measures of access, equity, quality, efficiency, and healthy lives (Davis, Stremikis, Squires, & Schoen, 2014). As a share of the Gross Domestic Product (GDP), healthcare spending accounted for 17.1% in 2013 (Squires & Anderson, 2015). According to the National Health Expenditure (NHE) Data, in 2014, the US healthcare spending increased 5.3 percent following growth of 2.9 in 2013 to reach $3 trillion, or $9,523 per person. The faster growth in 2014 was primarily due to major coverage expansions under Affordable Care Act (ACA), particularly, for Medicaid and private health insurance.

![Figure 1.1. United States Healthcare Spending as a Share of GDP (Source: The Economist Intelligence Unit)](image)

NHE grew 5.8% to $3.2 trillion in 2015 and accounted for 17.8% of GDP. Medicare and Medicaid spending grew 4.5% and 9.7% respectively in 2015. Moreover, prescription drug spending increased 9% and hospital expenditure grew 5.6% in 2015 (Centers for Medicare & Medicaid Services, 2017).

Healthcare high costs can be attributed to clinical, demographic, geographic, economic and other factors. Due to the high spending in healthcare that is continuing to take almost 17% of USA GDP over the past years, a healthcare system reform is driven by the motive to change or develop policies to enhance the healthcare delivery system and control cost. The ACA initiated the Hospital Readmissions Reduction Program (HRRP), which requires CMS to reduce payments to Inpatient Prospective Payment System (IPPS) hospitals with excess readmissions for applicable chronic conditions with high readmission rates such as heart failure, pneumonia and chronic obstructive pulmonary disease (COPD) (Centers for Medicare & Medicaid Services, 2016).
Asthma is a leading cause of hospitalization among children, and about 15-50% of pediatric patients are readmitted after an index admission (Chung, Hathaway, & Lew, 2015).

Asthma is one of America’s most common and costly chronic diseases. According to the Centers for Disease Control (CDC) about 25 million Americans have asthma. Asthma places significant health and economic burden on patients, their families and society (Zein et al., 2016). It is the leading chronic disease in children and it’s the top reason for missed school days (Centers for Disease Control and Prevention, 2012). Asthma is associated with the utilization of health services and it causes almost 2 million emergency room visits each year (United States Environmental Protection Agency, 2013).

In 2017, the U.S. Center for Disease Control and Prevention reported that more than 14 million doctor visits and 439,000 hospital stays each year, it also reported that in 2015, 3,615 people died from asthma. Many of these deaths and services utilization are considered avoidable with proper treatment and care (Centers for Disease Control and Prevention, 2017).

The annual total cost of asthma is $56 billion (Centers for Disease Control and Prevention, 2012). Total cost includes direct cost of health services utilization (prescription medications account for more than 50% of direct cost of asthma) and indirect cost of missed school or work days and lost pay from illness or death (Barnett & Nurmagambetov, 2011). Several medical interventions have been used to control healthcare cost for asthma patients. Case management is one type of intervention that has proven to be successful, case managers role is to help asthma patients identify and address their challenge with asthma management. A study found that a follow-up call by case managers to caregivers of children with asthma have increased visits to asthma clinics by almost one-third and improved patient care practices (Fisher-Owens, Boddupalli, & Thyne, 2011). Also, another study reported a 33% drop in hospitalization rates and a 52% drop in length of hospital stay among adults with asthma after introducing a multidisciplinary team for case management (Burke et al., 2016).

In order to address the main issue of potentially avoidable high utilization of health services, better manage chronic diseases to improve the dynamics of the US healthcare systems (e.g. improve healthcare outcome and reduce cost) and to move up in the world rankings, there is a need to identify factors associated with high-risk patients and to invest in care management programs targeted at the right population. This study aimed to identify key factors of high-risk asthma patients to be enrolled in case management programs based on data analysis outcomes. This study analyzed patients’ claims data provided by a Louisiana based insurance company to identify key factors (clinical and non-clinical) that contribute to increased cost and frequent use of healthcare resources. The study population was asthma patients in Louisiana who are Medicaid insured and who have a primary or secondary asthma diagnosis. Moreover, this study explored if there is a true effect of case management program on reducing the total asthma related cost.
1.1 Study motivation

Although there is no cure for asthma, it can still be managed with proper treatment and prevention care. Proper asthma management particularly for high-risk patients can aid in improving the quality of life for patients with asthma and in controlling the cost associated with health resources utilization. High risk patients are those patients that are at risk of not managing their condition, not adhering to their medication and not following doctors’ orders, which will result in adverse health outcomes. A study of a random sample of 1678 adults in Southeast Virginia found that those with asthma were more likely to report poor physical health and utilize more treatment services such as hospitalization and emergency department visits as compared to those without asthma (Behr, Diaz, & Akpinar-Elci, 2016). Consequently, asthma patients are characterized as frequent users of primary health services and represent a targeted population in care management programs. A successful investment in case management programs targeted at the right population of asthma patients may result in better practice of case management programs, improved health outcomes, less demand on health services and decreased cost.

1.2 Research objectives

The main goal of this study was to identify key factors associated with frequent use of primary health services and increased cost via asthma patients’ claims data. Moreover, this study explored the effectiveness of case management programs in reducing utilization and cost, and, finally, this study built two prediction models for total cost and compared the accuracy of prediction between the two models: linear regression model and Neural Networks and proposed the threshold point in which the Neural Networks model outperforms the linear regression in the accuracy of prediction for total cost.
Chapter 2. Literature Review

2.1 Asthma definition and statistics

The National Institutes of Health (NIH) defines asthma as a chronic lung disease that inflames and narrows the airways. It restricts the passage of air into the lungs and leads to episodes of wheezing, coughing, chest tightness, and shortness of breath. According to the World Health Organization (WHO), asthma is the most common disease among children and currently, an estimated of 235 million people suffer from asthma world-wide. Asthmatic patients experience episodic flare-ups (exacerbations) that might be acute or subacute. Severe asthmatic episodes can close off airways completely and, in some cases, may be life-threatening. When a patient experiences increased symptoms and deterioration in lung functions, an intense treatment will be required (Global Initiative for Asthma - GINA, 2017). Patients can get asthma attacks if they are exposed to “asthma triggers” which may result in costly utilization of healthcare services such as emergency department visits and hospitalization. According to the Centers for Disease Control and Prevention, the most common triggers that may cause asthma attacks are: tobacco smoke, dust mites, outdoor air pollution, pets, smoke from burning wood or grass, mold and other infections linked to flu or cold. In 2015, 46.9% of asthma patients reported having one or more asthma attacks (Centers for Disease Control and Prevention, 2017).

Asthma prevalence has been receiving a lot of attention by researchers. According to the Centers for Disease Control and Prevention summary health statistics, 7.6% of adults in the United States currently have asthma compared to 8.4% of children (Centers for Disease Control and Prevention, 2015). Asthma prevalence refers to percentage of people who have ever been diagnosed with asthma and still have asthma and it’s been increasing, the percentage of people diagnosed with asthma in the United states increased from 7.3% in 2001 to 8.4% in 2010. Moreover, an estimated 25.7 million people had asthma: 18.7 million adults aged 18 and over, and 7.0 million children aged 0–17 years as reported in 2010 (Akinbami et al., 2012).

In 2008, an estimated 264,428 adults and 97,069 children in Louisiana had asthma (Centers for Disease Control and Prevention, 2008). Asthma prevalence rate for Medicaid recipients in 2008 was 2.2% and the Medicaid population aged 10 years or less had the highest number of enrollees in 2007 and 2008 with a total number of claims 225,384 in both 2007 and 2008 (Asthma Management and Prevention Program & Louisiana Department of Health & Hospitals, 2010).

In 2011, The Louisiana Department of Health and Hospitals issued a Louisiana Asthma Burden Fact sheet that reported asthma prevalence based on demographics and economic characteristics. It was found that females are more likely to have the disease than males, and blacks are more likely to have asthma as compared to whites. Adults with an income of less than $15,000 and those with education less than high school have significantly higher rates of asthma than those in higher income and education levels (Louisiana Department of Health & Hospitals, 2011).
An analysis of the Louisiana Hospital Inpatient Discharge Database found a total of 21,398 adult inpatient hospitalizations during the period 2006 – 2011, with asthma as the primary discharge diagnosis (Lewis, Lackovic, & Soileau, 2015). During 2002-2008, children under five had the highest hospital discharge rates of all age groups, and the rate for asthma hospitalizations decreased during adolescent years, but increased around age 30 (Asthma Management and Prevention Program: Louisiana Department of Health and Hospitals, 2008). About 23 percent of high school students have missed at least one or more school days due to asthma (Louisiana Department of Health & Hospitals, 2011). It is thus obvious that asthma places a burden of the economy and society in Louisiana.

2.2 Health resources utilization statistics and cost

Health services utilization is defined as the quantity of health services used by patients. Health services/resources include but are not limited to: hospitalizations, readmissions, emergency department visits, urgent care clinics visits, outpatient office visits and prescription drugs. Hospital readmission is defined by the Centers for Medicare and Medicaid Services (CMS), as a patient admission to a hospital within 30 days of discharge from the same or another hospital (Centers for Medicare & Medicaid Services, 2016). Nearly twenty percent of Medicare patients discharged from hospitals are readmitted within 30 days (Jencks, Williams, & Coleman, 2009a). The Medicare Payment Advisory Commission (MedPAC) had estimated that 12% of readmissions are potentially avoidable with improved discharge process in terms of appropriateness of discharge timing and site (home or other care facilities). Preventable hospital readmissions contribute to a big part of avoidable medical spending. According to the data from the Center of Health Information and Analysis (CHIA) the estimated annual cost of hospital readmission for Medicare patients is $26 billion yearly with approximately $17 billion considered preventable with careful planning and communication among patients’ providers, caregivers, community social services and patients’ themselves (Reardon, 2015). Consequently, one of the key aspects of healthcare reform in the US is the attempt to lower the number of avoidable hospitalizations and other health services utilization as a way to improve quality of care and reduce cost.

As mentioned before, the annual cost for asthma is about $56 billion (United States Environmental Protection Agency, 2013). Direct costs make up almost $50.1 billion while indirect costs make up $5.9 billion (Barnett & Nurmagambetov, 2011). Direct costs include inpatient care, emergency department visits, ICU admissions, urgent care clinics visits, physician services, nursing services, ambulance use, drugs, devices and outpatient diagnostic tests, while the indirect costs include lost pay from sickness, lost work output from missed school and work days and asthma-related death (Krahn, Berka, Langlois, & Detsky, 1996).

In the US, asthma is the single most common condition in children and one of the leading causes of hospital admission and readmissions, In 2010, asthma accounted for 3,404 deaths and 439,400 hospitalizations (Centers for Disease Control and Prevention, 2013). In 2013, 6.5% of visits to office-based physician indicated asthma on the medical record, and a total of 1.6 million
visits to emergency departments with asthma as the primary diagnosis (Centers for Disease Control and Prevention, 2013b). According to the National Hospital Ambulatory Medical Care Survey in 2010, the number of hospital outpatient department visits was 1.3 million for asthmatic patients (Centers for Disease Control and Prevention, 2010).

Previous studies had investigated the healthcare costs and resource utilization associated with asthmatic patients and found that the average cost per asthma-related hospital stay for children remained relatively stable at about $3,600 from 2000 to 2010, whereas the average cost per asthma-related hospital stay for adults increased from $5,200 to $6,600 (Barrett, Wier, & Washington, 2014). A nationwide, cross-sectional study using data from the Medical Expenditure Panel survey found that the average charge for an asthma-related outpatient ED visit from 2006 to 2008 was $1,502 (Wang, Srebotnjak, Brownell, & Hsia, 2014).

Since 2009, the cost of inhaled asthma medicines have increase by an average of 50 percent (Consumer Reports, 2013). In 2014, Medicaid spent about $67 per member each year on asthma medicine, which is the third highest of any category. However, by the end of 2014, the price of asthma medications per member each year has dropped by almost 15 percent mostly because the cost per unit has dropped, this in turn has caused asthma to drop to the seventh most costly illness (The Express Scripts Lab, 2015). Moreover, it’s worth mentioning that an estimated 54.9 percent of adults and 78.3 percent of children are not committed to using the medications (Consumer Reports, 2013).

2.3 Factors associated with higher health resources utilization

There are a variety of clinical and non-clinical factors that can influence the level of patients utilization of primary health services. Clinical factors include poor-quality care, insufficient care coordination between providers (physicians, nurses and pharmacists), patients illness severity, and any adverse outcomes. It’s also worth mentioning that increased cost and utilization can be largely driven by patients circumstances and behavior, such as lack of social support and patients adherence to treatment, which are out of the healthcare providers control. Therefore, previous studies explored non-clinical factors (i.e. demographic and socio-economic) influence on readmission rates and other primary care services.

The diagnosis of asthma severity level has been investigated in the past as a potential reason for higher health resources utilization and increased cost. According to GINA (Global Initiative for Asthma - GINA, 2017), asthma severity is assessed retrospectively from the level of treatment required to control symptoms and exacerbations.

- **Mild asthma** is asthma that is well controlled with step 1 or step 2 treatment (as-needed reliever medication alone, or with low intensity controller treatment such as low dose ICS[^1^], leukotriene receptor antagonists or chromones).

[^1^]: ICS: long-term control medications, used for prevention/control of asthma, not treatment of acute exacerbations
• **Moderate asthma** is the asthma that is well controlled with low dose ICS/LABA\(^2\).

• **Severe asthma** is the asthma that requires a high-dose ICS/LABA, to prevent it from becoming uncontrolled, or asthma that remains uncontrolled despite this treatment.

It’s important to distinguish between asthma severity and asthma control. The GINA guidelines define asthma control as “the extent to which the effect of asthma can be seen in the patient or have been reduced or removed by treatment”. Asthma control has two aspects: symptom control and risk factors for poor future outcomes. Poor symptom control is a burden to patient and a risk factor for flare-ups. A risk factor is a factor that increases the patients’ future risk of exacerbations, loss of lung function, or medication side effect. (Global Initiative for Asthma - GINA, 2017).

### 2.4 Intervention programs

A medical intervention in general is defined as any medical procedure or application that is intended to relieve an illness or injury. For asthma patients, an intervention aims to help patients better control their disease and prevent asthma exacerbations, which can reduce the economic burden of asthma and improve patients quality of life. Although the exact cause of asthma is unknown and it cannot be cured, it still can be controlled with self-management education, appropriate medical care, and avoiding exposure to environmental triggers that might lead to asthmatic episodes. Uncontrolled asthma can be due to a number of factors with poor knowledge and lack of education about the subject being two of them. Several studies reported that about half of asthma patients don’t use the inhaler correctly (Al-Zahrani et al., 2015)(Consumer Reports, 2013). Thus, sufficient education of asthma control can aid in improving patients’ health outcomes and reduce the potentially avoidable ED visits and hospitalizations with better care.

#### 2.4.1 Home-based intervention

“The Community Preventive Services Task Force (CPSTF) recommends the use of home-based multi-trigger, multicomponent interventions with an environmental focus for children and adolescents with asthma to improve overall quality of life and productivity, specifically: improving asthma symptoms and reducing the number of school days missed due to asthma. Interventions involve home visits by trained personnel and aim to reduce exposure to multiple indoor asthma triggers (allergens and irritants) through 2 or more activities. Activities may include: assessment of the home environment; changing the indoor home environment to reduce exposure to asthma triggers; or education about the home environment. Most programs also include non-environmental activities such as training to improve self-management,

---

\(^2\) ICS/LABA combination in a single inhaler represent safe, effective and convenient treatment options for the management of patients with asthma and COPD
education, social services and support, or coordinated care” (The Community Guide & Centers for Disease Control and Prevention, 2014).

A systematic review of 20 studies on home-based interventions with environmental focus to reduce asthma-related outcomes reported reduced asthma symptom-days by 0.8 days per 2 weeks, reduced missed school days by about 12 days per year and reduced asthma acute care visits by 0.57 visits per year (Crocker et al., 2011).

2.4.2 Asthma self-management education

Self-management education is one of the strategies that the CDC’s National Asthma Control Program (NACP) suggest to implement to educate patients with asthma on how to engage in behaviors and practices that help them gain optimal control of their disease. Self-management is usually guided by a healthcare professional and it includes asthma education, self-monitoring of symptoms and lung function as well as adjusting treatment according to the action plan provided by a healthcare provider (Van der Meer et al., 2011). An analysis of a US claim database found that 32% of asthma patients who stepped down their medication had an asthma exacerbation in the two years following the step-down event (Rank et al., 2015). It was found that families who received an individualized asthma education during patient’s hospitalization by trained volunteers had better asthma management behaviors compared to patients who received the standard medical management (Rice et al., 2015). A study investigated the effectiveness of different educational programs in improving asthma control and it was found that individual verbal instructions and integrated asthma classes showed greater improvement in QOL questionnaire and asthma control compared to written information in an asthma booklet (Urek et al., 2005).

2.4.3 Multidisciplinary interventions

In their study, Dinelli & Higgins reported that a combined intervention that includes patient education, a coordinated self-monitoring plan and patient follow-up was associated with improved care and cost-saving, particularly in reducing the number of clinic visits, the number of ordered chest radiographs and asthma relief medications (Dinelli & Higgins, 2002).

Another study found that a multidisciplinary case management strategy that consisted of a consultant, specialist nurse, physiotherapist and psychologist was effective in reducing hospitalization rates for patients with persistent asthma (Burke et al., 2016).

2.4.4 Disease and case management

Disease management (DM) is a coordinated healthcare system that educates patients with chronic diseases on how to better manage their disease and prevent complications. The goal of DM is to identify persons at risk for chronic conditions, to promote self-management by patients or their caregivers, and to help achieve the best clinical outcomes (Schrijvers, 2009). Systematic review and meta-analyses found considerable evidence in the effectiveness and
efficiency of DM programs in improving the adherence to guidelines, health outcomes and quality of life among patients with chronic illnesses (Hisashige, 2013).

Asthma is one of the chronic conditions that have been the focus of DM programs. A recent study evaluated the effectiveness of disease management programs for adults with asthma, patients who were enrolled in chronic disease management program reported improvements in asthma-specific quality of life, severity scores, and lung function tests compared to usual care (Peytremann-Bridevaux, Arditi, Gex, Bridevaux, & Burnand, 2015). Another study have reported an accumulation of circumstantial evidence that DM programs targeted at asthma patients have reduced the utilization of health resources and have improved the aspects of self-management and organization of care after the implementation of the program (Steuten, Lemmens, & Vrijhoef, 2007).

According to Case Management (CM) Society of America, “case management is a collaborative process of assessment, planning, facilitation, care coordination, evaluation, and advocacy for options and services to meet an individual’s and family’s comprehensive health needs through communication and available resources to promote quality cost-effective outcomes”.

Case management is distinguished from disease management in that it targets high-risk patients who due to diverse combinations of health, social and functional problems, are likely to need hospitalization. Disease management targets patients who have one major diagnosis and who, because of their major diagnosis, have a relatively standard set of needs. Also, Case management is often episodic, the interventions are needed over a specific period of time or until certain goals are met, or a patient transitioned from one level of care to another. Disease management tend to be long term and goes beyond the short-term nature of a health encounter (Ahmed, 2016).

State Medicaid care management programs considers a process for enrollment in the programs. First, States must select the eligible population that the program will target, then determine how they will identify potential members and finally decide the enrollment strategy based on the program. According to Exhibit 3.1. Care management population selection and enrollment process issued by the Agency for Healthcare Research and Quality, the Medicaid disease and care management programs select eligible population among specific disease, high-risk members and population-based approach. Care management programs can target specific chronic conditions or focus on high-risk and high-cost beneficiaries. These programs can also follow a "population-based" approach by including their entire fee-for-service (FFS) and primary care case management (PCCM) population or Temporary Assistance for Needy Families (TANF) and Supplemental Security Income (SSI) populations within FFS and PCCM and provide the interventions appropriate for the member's risk level or disease. After selecting the eligible population, these programs must identify and stratify members by utilizing several tools such as claims data, risk stratification, health assessment tools and predictive modeling. Finally, the
technique in which the program chooses to enroll and engage members varies based on program
design (Agency for Healthcare Research and Quality, Rockville, MD, 2014).

2.4.5 Common success factors for case management from literature

A previous study identified asthma interventions from previous literature that had
positive asthma outcomes, and identified the following characteristics to be associated with a
successful asthma program “1. were community based, (2) engaged the participation of
community-based organizations, (3) provided program components in a clinical setting, (4)
provided asthma training to health-care providers, (5) collaborated with other organizations and
institutions and with government agencies, (6) designed a program for a specific racial/ethnic
group, (7) tailored content or delivery based on individual health or educational needs, and (8)
conducted environmental assessments and tailored interventions based on these assessments”
(Clark, Lachance, Milanovich, Stoll, & Awad, 2009a).

Another study that reviewed asthma programs targeted at children with asthma found
that the common factors for success involve case managers spending time contacting and
patiently and persistently working with the family, which contributes to building a trusting
relationship, consequently, the positive outcomes benefit all parties involved (Schulte, Musolf,
Meurer, Cohn, & Kelly, 2004)

Developing and implementing an intervention program targeted at asthma patients with
the previously mentioned characteristics will be more likely to yield positive outcomes.

2.5 High-risk patients definition from previous studies

CMS contracted the National Committee for Quality Assurance (NCQA) to develop a tool
to evaluate quality and performance on important dimensions of care and service, this tool is
referred to as The Health Plan Employer Data and Information Set (HEDIS) (Centers for Medicare
and Medicaid Services, 2017).

HEDIS developed a definition of patients with asthma who are at risk of adverse outcomes
by analyzing 12-months of utilization data. Patients who had at least one or more of the following
types of utilization were defined as high-risk:

1. At least one hospitalization in the past year with an asthma diagnosis and had at least
one asthma prescription during that year
2. At least one emergency department or urgent care visit in the past year with an
asthma diagnosis and had at least one asthma prescription during that year
3. At least two office visits in the past year with an asthma diagnosis and had at least
one asthma prescription during that year
4. At least one visit in the past year with an asthma diagnosis and another in the past 18
months and had at least one asthma prescription
5. At least four prescriptions for asthma medication in the past year (Bennett, Lozano, Richardson, McCauley, & Katon, 2008)

The Global Initiative for Asthma (GINA) guidelines for identifying risk factors for poor asthma outcomes states that having one or more of the following risk factors increases the risk of exacerbations even if asthma is well controlled:

- Uncontrolled asthma symptoms
- ICS not prescribed, poor ICS adherence; incorrect inhaler technique
- High SABA\textsuperscript{3} use
- Low FEV
- Major psychological or socioeconomic problems
- Exposures: smoking, allergen exposure
- Comorbidities: obesity, rhinosinusitis; confirmed food allergy
- Sputum or blood eosinophilia; elevated FENO in allergic adults
- Pregnancy
- Ever been intubated or in intensive care for asthma
- Having 1 or more severe exacerbations in the last 12 months

Previous studies have identified high-risk patients who will potentially over-utilize the primary health services in terms of their demographics, socio-economic status, disease severity and level of utilization of primary health services. There were multiple data sources including hospital databases, claims data, interviews and questionnaires which we will explore in the next section.

2.6 Previous studies data sources

2.6.1 Hospital databases

A study in Thailand that used hospital electronic databases to analyze asthmatic patients data, reported that patients diagnosed with high severe asthma had higher utilization of healthcare resources compared to patients who had mild/moderate severe asthma (Dilokthornsakul, Lee, Dhippayom, Jeanpeerapong, & Chaiyakunapruk, 2016).

Nath & Hsia utilized data from the National Hospital Ambulatory Medical Care Survey and found that Children who were younger, male, belonged to racial or ethnic minorities, and insured with Medicaid/CHIP had higher ER visits (Nath & Hsia, 2015).

In their study that utilized Public Health Service Hospitals database in Spain, Gonzalez-Barcala et al. found that female patients over 60 years of age had higher hospitalization rates due to asthma and longer hospital stays (Gonzalez-Barcala et al., 2011).

\textsuperscript{3}SABA: Short-Acting Beta Agonists are rescue medications that are fast acting and temporarily relief symptoms
2.6.2 Government agency database

Another study analyzed data collected by the Agency of Healthcare Research and Quality through the Medical Expenditure Panel Survey of children aged from 0-11 years during 2001-2006 period. The study found that exposure to asthma triggers such as secondhand smoke was associated with higher utilization of hospitals and emergency departments among children with asthma as compared to children without asthma (Jin, Seiber, & Ferketich, 2013).

2.6.3 Interview and questionnaires

In their study, Eisner et. al found that non-white race, lower income, and greater asthma severity level were identified as risk factors for subsequent hospitalization among adults with asthma (Eisner, Katz, Yelin, Shiboski, & Blanc, 2001).

A previous study that conducted interviews with minority inner city asthma patients and collected data on their ED visits and hospital admissions, the study reported that patients with no established primary care provider, spoke mostly Spanish and reported allergy to cockroaches were associated with higher resource utilization and worse quality of life (Wisnivesky, Leventhal, & Halm, 2005).

An administered a questionnaire in which a random, stratified sample of 1678 adult were asked to self-report their primary care visits, emergency department visit, admission to a hospital found that those diagnosed with asthma were more likely to frequently utilize primary care services for matters associated with management of condition, treatment or hospitalization after an acute and episodic attack (Behr et al., 2016).

2.6.4 Insurance claims data

In Germany, an analysis of claims data found that patients with a primary or secondary asthma diagnosis had significantly higher inpatient, outpatient visits as well as a higher number of pharmaceuticals prescriptions compared to patients without asthma. Moreover, the study also showed that patients with persistent asthma had higher asthma-related cost compared to patients with intermittent asthma (Jacob et al., 2016).

2.7 Research gap

Previous studies reported asthma prevalence (Lewis et al., 2015), (Akinbami et al., 2012), statistics and cost of utilizing primary health services (Reardon, 2015), (Jencks, Williams, & Coleman, 2009b), (Krahn et al., 1996), (Wang et al., 2014). Other studies looked at different types of intervention programs targeted at asthma patients and reported the outcomes in terms of better care and lower cost (Crocker et al., 2011), (Van der Meer et al., 2011), (Urek et al., 2005), (Burke et al., 2016) and identified common success factors for asthma programs with positive outcomes (Clark, Lachance, Milanovich, Stoll, & Awad, 2009b), (Schulte et al., 2004).
Moreover, the literature had previous work that defined high-risk patients that are more likely to overutilize primary health services while utilizing multiple database sources such as: government (Jin et al., 2013), hospital databases (Dilokthornsakul et al., 2016), (Nath & Hsia, 2015), interview and surveys (Eisner et al., 2001), (Wisnivesky et al., 2005), (Behr et al., 2016), and insurance claims (Jacob et al., 2016).

There is a need for a comprehensive study that utilizes patients’ claims data and propose a holistic definition of high-risk patients based on their demographics, history of utilization of health care services and cost. This study utilized Medicaid patients’ claims data in from an insurance agency in Louisiana to identify factors associated with high-risk asthma patients that could benefit from a case management program, moreover, this study also compared the accuracy of prediction between a linear regression model and neural networks, and presented the number of data points needed to train neural network to outperform regression.

In the health insurance agency where this study took place, all patients with an asthma diagnosis or any other chronic illness are considered eligible for case management. However, a common observation among patients in the eligible population is that they have low enrollment rate in case management (for asthma, enrollment rate is 14% among all members). That is, case managers would reach out to patients and offer them the service, and the patients have the choice to accept or decline the enrollment in case management. This project analyzed claims data to identify, if any, factors are associated with frequent use of health services and increased cost among Medicaid asthmatic patients. The higher risk patients can constitute a population that might require a particular approach to improve the chances of a patient accepting to be enrolled in case management.
Chapter 3. Research Methodology

3.1 Study objective

The primary objective of this study was to identify asthma patients’ key factors associated with frequent use of primary health services and increased cost via claims data. Moreover, this work explored the effectiveness of case management program in reducing total asthma related cost.

In addition to the primary objective, this study compared the accuracy of two models predicting patients’ total cost: linear regression model vs neural networks (NN), and presented a heuristic case study for finding the optimal number of data points needed to train the neural network to outperform linear regression model in terms of accuracy of prediction.

3.2 Study setting

The data source in this study was a healthcare insurance provider based in Baton Rouge, Louisiana. The company provides Medicaid or LaChip for qualified members through state’s Healthy Louisiana Program and links Medicaid insured members to primary care providers, pharmacies and case managers.

3.3 Ethical consideration

Institutional Review Board at Louisiana State University reviewed and approved this study. The researcher received HIPPA training at the healthcare insurance provider site prior to the beginning of the study. Patients’ information was de-identified by the health insurance company’s data analytics team according to HIPPA and company policy guidelines before sharing the data with the researcher. A generic unique ID was created for each member to link patient information across claims files anonymously to protect patients’ information.

3.4 Study population inclusion and exclusion criteria

The study population was Medicaid insured patients of all ages who have a primary or secondary asthma diagnosis and have had a record of Medicaid Insurance during the study period January 1st, 2015 to November 30th, 2017. Members with asthma were identified by using International Classification of Diseases. The ninth revision codes (ICD-9Dx) with 493.XXX asthma codes and the tenth revision codes (ICD-10Dx) with J45.XXX as asthma codes. Drug NDC\(^4\) number was used to identify asthma medications. Patients were eligible to be included in the study if they have the following criteria:

- Patients who had a record of Medicaid insurance during the study period

\(^4\) National Drug Code: A unique 10-digit, 3-segment number. It is a universal product identifier for human drugs in the United States
Patients with primary or secondary diagnosis of asthma
- Patients who had at least one primary care claim (physician office visit) during the study period with asthma as the primary service diagnosis

Patients who were deceased during the study period, patients who had no primary care claim due to asthma during the study period, patients who are older than 75 and patients who had missing or invalid information were excluded from the study. Based on the inclusion and exclusion criteria 12,029 patients were in the study and 127,762 patients were excluded from the study. Patients who were older than 75 years of age were excluded from the study to comply with HIPPA regulations to reduce the probability of identifying the subject.

3.5 Data collection

Figure 3.1 displays the process of data collection process in terms of data categories extracted. Data was extracted from MicroStrategy database, each data category was combined in separate excel sheets, patients’ information were linked across data categories using patients’ ID. To protect patients’ information, patients ID were de-identified into unique generic unique ID that was used in the excel sheets received by the researcher.
The details of the data categories collected from each claim source are displayed in Figure 3.2. Member data includes age, gender, race, zip code was used to find median income and utilized as a socio-economic indicator and patient enrollment CM (Yes/No) and the date of enrollment in CM. Hospital & ICU admissions claims data included admission diagnosis, admission date, charge data, length of hospital stay, and billed amount. Billed amount refers to the total amount that was billed for the cost not just the amount paid by insurance provider. Paid Amisys amount is the amount covered by the insurance provider. Pharmacy data included drug description, fill date, drug NDC number, billed amount and paid Amisys amount. Emergency Department, Urgent Clinic and Primary care visits data included claim’s primary Diagnosis Service description, service date, billed amount and paid Amisys amount.
Figure 3.2. Collected Data: Linked by generic unique ID

Data was cleaned and filtered for the study population based on the inclusion criteria. Admission Diagnosis was used to filter asthma related claims for hospital and ICU admissions claims. Primary Diagnosis Service was used to filter asthma related claims in Emergency Department, Urgent Care Clinics and Primary care claims.

For pharmacy files, asthma medications NDC codes were used to identify asthma medications claims. Patients’ claims with a primary service diagnosis other than asthma was not included in this study.

A simple R-based script was written to clean the dataset from incomplete, false, inaccurate and missing information. Patients with missing demographic information, or socio-economic indicators (zip code) were removed from the dataset. Furthermore, patients with extreme billed amount values corresponding to a single claim were removed from the dataset. Patients who had zero billed amount corresponding to a claim were also removed. The dataset then was conformed to standard notations, features and numeric value presentations.

3.6 Study population descriptive statistics

The sample consists of 9,977 asthma patients from all ages including infants and adults (0 to 73 years). The sample has 4,651 females and 5,326 males. A total of 1,609 patients are enrolled in case management, which translates to approximately 16% enrollment rate.

Table 3.1 displays the breakdown of the study population by age groups; Note that the majority of the population is aged 18 and under.

Table 3.1. Population age groups breakdown

<table>
<thead>
<tr>
<th>Age Range for study population</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-18</td>
<td>8,409</td>
</tr>
<tr>
<td>19-37</td>
<td>835</td>
</tr>
<tr>
<td>38-56</td>
<td>498</td>
</tr>
<tr>
<td>57-73</td>
<td>235</td>
</tr>
<tr>
<td>Total</td>
<td>9,977</td>
</tr>
</tbody>
</table>
The PROC MEANS by SAS was used to provide descriptive statistics for total cost, age, and median income.

As shown in Table 3.2 the mean of the total cost for the study population is $240 with a minimum of $3 and a maximum close to a $1,000. The standard deviation indicates that there is no large dispersion of data around the mean, however, by looking at the maximum value, we expect to find outliers in the data.

Table 3.2. Descriptive Statistics for Total Cost

<table>
<thead>
<tr>
<th>Analysis Variable: Total_Cost</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9977</td>
<td>240.4</td>
<td>221.9</td>
<td>3.0</td>
<td>999.8</td>
</tr>
</tbody>
</table>

Table 3.3 presents the descriptive statistics for variable age; the mean is 13 years and the standard deviation equals 12.8, which is consistent with what we mentioned earlier about the majority of the study population being aged 18 and under. Moreover, the maximum is 73, technicality introduced to comply with HIPPA regulations.

Table 3.3. Descriptive Statistics for Age

<table>
<thead>
<tr>
<th>Analysis Variable: Age</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9977</td>
<td>13.1</td>
<td>12.8</td>
<td>0</td>
<td>73</td>
</tr>
</tbody>
</table>

The median income was collected by utilizing the patients’ zip codes. Table 3.4 displays the descriptive statistics for median income. The mean income is about $42,000 and the standard deviation is $6,966.

Table 3.4. Descriptive Statistics for Income

<table>
<thead>
<tr>
<th>Analysis Variable: Income</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9977</td>
<td>42082.0</td>
<td>6966.2</td>
<td>24038.0</td>
<td>109257.0</td>
</tr>
</tbody>
</table>

Since there is a low enrollment rate in CM, we also added the descriptive statistics of patients’ demographics and income for those patients who chose to be enrolled in CM.

Table 3.5 displays the count for each gender for patients in enrolled in CM. it shows an approximately equal ratio between male and female patients.

Table 3.5. Gender count for patients enrolled in CM

<table>
<thead>
<tr>
<th>Male CM patients count</th>
<th>Female CM patients count</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>823</td>
<td>786</td>
<td>1609</td>
</tr>
</tbody>
</table>
Table 3.6 displays the descriptive statistics for age variable. The mean age for patients who are case managed is 17 with a low standard deviation of 0.40; this indicates that the majority of patients enrolled in CM are younger patients.

**Table 3.6.** Descriptive statistics for Age for patients enrolled in CM

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1609</td>
<td>17</td>
<td>0.40</td>
<td>0</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 3.7 presents the descriptive statistics for income variable. The mean income for enrolled patients is about $41,750 and the standard deviation is 7696.

**Table 3.7.** Descriptive statistics for Income for patients enrolled in CM

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1609</td>
<td>41758.9</td>
<td>6796.0</td>
<td>24038</td>
<td>63716</td>
</tr>
</tbody>
</table>
Chapter 4. Results

4.1 Difference in cost between patients who are case managed vs non-case managed

The SAS statistical software was used to analyze the difference in total cost between patients who are case managed vs. patients that are not case managed. In order to utilize a T-test comparison, we started by checking the appropriate conditions, using a standard pre-test analysis. By design we can guarantee independence of samples between and within the two groups. In the absence of a true sigma for standard error in the variable Total Cost the pooled sigma is assumed to be a reliable estimate.

First of all, we tested the Total Cost’s normality assumption. In other words, we performed a goodness of fit test with the following hypothesis:

\[ H_0 : \text{"Total Cost" follows a normal distribution} \]
\[ H_1 : \text{"Total Cost" doesn’t follow a normal distribution} \]

As shown Table 4.1, the Goodness of fit tests all have p-value smaller than 0.05, therefore, we reject the null hypothesis that the data is normally distributed. Moreover, as displayed in Figure 4.1, the histogram shows that the distribution is skewed to the right.

<table>
<thead>
<tr>
<th>Goodness-of-Fit Tests for Normal Distribution</th>
<th>Statistic</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.174</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>109.1</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>616.6</td>
</tr>
</tbody>
</table>

Second, we performed a test for homogeneity of variance using Levene’s test for total cost calculated for gender and CM as they are the only two variables that can generate categories, and the following hypothesis was tested:

\[ H_0 : \text{All variances are equal} \]
\[ H_1 : \text{All variances are not equal} \]

As shown in Table 4.2, the p-value is low, so we reject the null hypothesis, which implies that variances are not equal.
Table 4.2. Levene's test for homogeneity of variance for Total Cost

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender*CM</td>
<td>3</td>
<td>2.54E11</td>
<td>8.467E10</td>
<td>9.80</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9973</td>
<td>8.616E13</td>
<td>8.6396E9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.1. Distribution of Total Cost data
Figure 4.2. Boxplots for the distribution of Total Cost for gender and CM levels

We noted the skewness of our data, which is actually a feature worth exploring on its own, but continued nonetheless without removing the outliers performing a T-test to compare the cost between the two sets of interest. The first set contains the patients who are enrolled in that case management program (designation 1) for at least a year, and the second those who are not (designation 0).

Our T-test hypothesis was formed based on our research question, thus our null and alternative hypothesis are:

\[ H_0: \text{Patients enrolled in CM or not, have the same average total cost} \]

\[ H_1: \text{Patients enrolled in CM have different average total cost} \]

Based on the two-sided T-test results displayed in Table 4.3, the computed p value is much smaller than 0.05, therefore, the null hypothesis that patients enrolled and patients not enrolled have the same total cost is rejected. By further inspecting the mean total cost for enrolled and not enrolled, we discover that patients enrolled in case management have a higher total cost on average. This may indicate, that since their cost was high, they were either asked to enroll or were self-enrolled in case management. However, this also raises a question about the effectiveness of the CM program at the insurance company where this study takes place, and prompted us to further investigate how the cost changes overtime for patients who are case managed vs non-case managed.
Table 4.3. T-test comparison between opting in or out from CM for total cost variable

<table>
<thead>
<tr>
<th></th>
<th>Method</th>
<th>Mean</th>
<th>95% CL Mean</th>
<th>Std Dev</th>
<th>95% CL Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Not Enrolled</td>
<td>232.1</td>
<td>227.4</td>
<td>236.7</td>
<td>217.1</td>
<td>213.9</td>
</tr>
<tr>
<td>1 Enrolled</td>
<td>285.4</td>
<td>255.7</td>
<td>315.2</td>
<td>240.9</td>
<td>221.7</td>
</tr>
<tr>
<td>Diff (1-2) Pooled</td>
<td>-53.3</td>
<td>-80.4</td>
<td>-26.1</td>
<td>217.9</td>
<td>214.7</td>
</tr>
<tr>
<td>Diff (1-2) Satterthwaite</td>
<td>-53.3</td>
<td>-83.4</td>
<td>-23.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Method     | Variances | DF    | t Value | Pr > |t| |
|------------|-----------|-------|---------|-------|---|
| Pooled     | Equal     | 8621  | -3.85   | 0.0001|
| Satterthwaite | Unequal   | 266.7 | -3.49   | 0.0006|
| Cochran    | Unequal   |       | -3.49   | 0.0006|

<table>
<thead>
<tr>
<th>Method</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folded F</td>
<td>254</td>
<td>8367</td>
<td>1.23</td>
<td>0.0158</td>
</tr>
</tbody>
</table>

Figure 4.3. Displays side by side boxplots for the two groups, patients enrolled in CM and patients not enrolled in CM. The range and the mean of Total Cost is higher for patients enrolled. The same figure shows the distribution of Total Cost variable for both groups, the top graph presents the patients not enrolled in case management and the bottom one those enrolled in it.

Figure 4.3. Side by side boxplot and histograms for enrolled and non-enrolled patients
4.2 Exploring the effect of case management on cost over time

To further investigate how enrollment in case management changes the total cost over time and to compare that to the total cost over time for patients who have never been enrolled in case management, two random samples were selected from the study population. Each sample has 60 patients with their total cost in 2015 and total cost in 2017; 30 patients have been enrolled in case management in 2016 and 30 patients have never been enrolled in case management.

The relative difference in total cost between 2017 and 2015 was found for the two samples of patients and a T-test procedure was performed to identify if there is a statistically significant difference in how the total cost changes over time based on patient’s enrollment in CM. The following equation was used to find the relative difference in total cost:

\[
\frac{\text{Total Cost 2015} - \text{Total Cost 2017}}{\text{Total Cost 2015}}
\]

And the following hypothesis was tested:

\[H_0: \text{The average relative difference between the cost in 2015 and 2017 is the same for the two groups}\]

\[H_1: \text{The average relative difference between the cost in 2015 and 2017 is not the same for the two groups}\]

Table 4.4 presents the T-test results for sample 1, since \( p \) is quite high we fail to reject the null hypothesis and thus there is no indication that there is a difference in the average relative difference cost changes between the two groups (CM vs NonCM).

**Table 4.4. Sample 1: T-test results**

<table>
<thead>
<tr>
<th>( t )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.61892</td>
<td>0.5453</td>
</tr>
</tbody>
</table>

Table 4.5 displays estimates for the mean of the relative difference between the two years of the two groups. We observe that in our sample the average relative difference of the non-case managed group is approximately -0.91 which implies an increase in the cost between 2015 and 2017.

Similarly, the average relative difference of the case managed group is approximately -2.6 which again implies an increase in the cost between 2015 and 2017 and it is higher than the CM group.
Similarly, for the second sample with another set of patients, we get the T-test results displayed in Table 4.6, and thus we once again fail to reject the null hypothesis.

Table 4.6. Sample 2: T-test results

<table>
<thead>
<tr>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.4721</td>
<td>0.1642</td>
</tr>
</tbody>
</table>

In other words, we have another indicator that suggests that there is no real change in enrolling in the case management service.

For this result, Table 4.7 displays the estimates for the mean of the relative difference between the two years of the two groups.

the average relative difference of the non-case managed group is approximately -3.05 which implies an increase in the cost between 2015 and 2017.

Similarly, the average relative difference of the case managed group is approximately -0.31 which again implies an increase in the cost between 2015 and 2017 and, in this case, it is higher than the non-case managed group.

Table 4.7. Sample 2: Estimates for average relative difference in the two groups

<table>
<thead>
<tr>
<th>Mean of difference in cost for non-case managed patients</th>
<th>Mean of difference in cost for case managed patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.0474568</td>
<td>-0.3107453</td>
</tr>
</tbody>
</table>

The analysis above indicates that on random samples in our dataset, the relative difference in cost is not immediately related to being case managed or not. Further analysis is needed on bigger samples and it will definitely be a priority in future work.

4.3 Forward variable selection procedure for total cost

Next, a forward variable selection procedure was used to identify variables that are statistically significant in a regression model with total cost as the dependent variable. Prior to that, we performed a multicollinearity detection test. Table 4.8 displays the result of multicollinearity diagenetic using Variance Inflation Factor (VIF). The variables that have VIF value above 5 have multicollinearity. Number of ER claims and ER billed amount, Rx prescription drugs, billed amount and Rx paid Amisys which is the amount covered by the insurance company, as well as urgent care clinics billed amount and paid Amisys amount all have multicollinearity.
Table 4.8. Multicollinearity diagnostic for regression input variables

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>CM</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>DurationCM</td>
</tr>
<tr>
<td>Income</td>
</tr>
<tr>
<td>ER</td>
</tr>
<tr>
<td>ER_Billed</td>
</tr>
<tr>
<td>ER_Amisys</td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Primary_Billed</td>
</tr>
<tr>
<td>Primary_Amisys</td>
</tr>
<tr>
<td>Rx</td>
</tr>
<tr>
<td>Rx_Billed</td>
</tr>
<tr>
<td>Rx_Amisys</td>
</tr>
<tr>
<td>Urgent</td>
</tr>
<tr>
<td>Urgent_Billed</td>
</tr>
<tr>
<td>Urgent_Amisys</td>
</tr>
</tbody>
</table>

The input variables for the forward variable selection are: Age, gender, enrollment in case management program, duration in enrollment in case management program, number of claims for hospital stays and length of stay, number of emergency department claims, number of asthma medication claims, number of primary care office visits, number of urgent care clinics claims, number of ICU admission claims, length of stay and the billed amount for each claim.

The analysis was carried over using the procedure PROC REG in SAS and the results are displayed in Table 4.9

Table 4.9. Summary of forward variable selection

<table>
<thead>
<tr>
<th>Summary of Forward Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>
The significant input variables according to our table are:

a) claims for primary physician office visits,
b) claims of emergency room visits,
c) claims of pharmacy prescriptions and
d) claims of urgent clinic visits.

The regression equation is

\[
\text{Total cost} = a_1 \times \text{PrimaryBilled} + a_2 \times \text{ERBilled} + a_3 \times \text{RxBilled} + a_4 \times \text{UrgentBilled} + b
\]

This conforms to our intuition since all these variables come with a cost attached to them. Other input variables such as gender and age do not have a statistically significant effect on total cost.

The effect of gender on total cost was explored by performing a T-test between the two groups using designation 0 for male and designation 1 for females. The following hypothesis was tested:

\[
H_0: \text{Male and female patients, have the same average total cost}
\]

\[
H_1: \text{Male and female patients, have different average total cost}
\]

Table 4.10 displays the T-test results. The p-value is lower than 0.05, therefore, we reject the null hypothesis. There is a clear difference between the average total cost with respect to the two genders in this dataset. On average male participants have a higher average total cost than that of females and their difference is statistically significant.

This result need be explored further, in an experiment where all over variables are controlled and the only difference between the two groups analyzed is the gender.

Figure 4.4 shows that the mean for male and female patients, where the mean for males is higher, and the solid black line indicates a big number of outliers in both groups.
Table 4.10. T-test for difference in Total Cost based on Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Method</th>
<th>Mean</th>
<th>95% CL Mean</th>
<th>Std Dev</th>
<th>95% CL Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Male</td>
<td>248.6</td>
<td>242.5</td>
<td>254.8</td>
<td>228.4</td>
</tr>
<tr>
<td>1</td>
<td>Female</td>
<td>231.1</td>
<td>224.9</td>
<td>237.2</td>
<td>213.9</td>
</tr>
<tr>
<td>Diff (1-2)</td>
<td>Pooled</td>
<td>17.5</td>
<td>8.84</td>
<td>26.2</td>
<td>221.7</td>
</tr>
<tr>
<td>Diff (1-2)</td>
<td>Satterthwaite</td>
<td>17.5</td>
<td>8.88</td>
<td>26.2</td>
<td></td>
</tr>
</tbody>
</table>

| Method   | Variances   | DF    | t Value | Pr > |t| |
|----------|-------------|-------|---------|------|---|
| Pooled   | Equal       | 9975  | 3.95    | <.0001 |
| Satterthwaite | Unequal | 9926.1 | 3.97    | <.0001 |
| Cochran  | Unequal     | .     | 3.97    | 0.0001 |

Equality of Variances

<table>
<thead>
<tr>
<th>Method</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folded F</td>
<td>5325</td>
<td>4650</td>
<td>1.14</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Figure 4.4. Side by side boxplots and histograms for male and female patients
4.4 Identifying factors associated with frequent use of health services

Proc GLM was used to identify which factors are associated with frequent use of primary health services.

The input variables are: Age, gender, income, and enrollment in case management program.

The output variables are: Number of claims for the primary health services (emergency room visits, primary care physician office visit, prescriptions, ICU visits and urgent clinic visits).

For the dependent variable ER (emergency room visits), the model is significant as the p-value is <0.0001 which is below 0.05 level of significance as displayed in Table 4.11.

Table 4.11. Regression model: Dependent variable ER vs. input variables

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>13</td>
<td>4.56</td>
<td>0.351</td>
<td>5.48</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9963</td>
<td>637.51</td>
<td>0.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9976</td>
<td>642.07</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Age and gender are not statistically significant in predicting the number of ER claims. The median income and enrollment in case management and their interaction are statistically significant.

For the variable primary (number of claims for doctor’s office visits), the result is displayed in Table 4.12. The regression model is significant with a p-value of <.0001 which is lower than the level of significance of 0.05 at 95% confidence level.
Table 4.12. Regression model: Dependent variable primary vs. input variables

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>13</td>
<td>152.26</td>
<td>11.71</td>
<td>4.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9963</td>
<td>29185.38</td>
<td>2.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9976</td>
<td>29337.65</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>1</td>
<td>0.34455308</td>
<td>0.34455308</td>
<td>0.12</td>
<td>0.7316</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>0.77082514</td>
<td>0.77082514</td>
<td>0.26</td>
<td>0.6080</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>0.00272967</td>
<td>0.00272967</td>
<td>0.00</td>
<td>0.9756</td>
</tr>
<tr>
<td>Income</td>
<td>1</td>
<td>0.53173249</td>
<td>0.53173249</td>
<td>0.18</td>
<td>0.6701</td>
</tr>
<tr>
<td>CM*Gender</td>
<td>1</td>
<td>0.32913689</td>
<td>0.32913689</td>
<td>0.11</td>
<td>0.7375</td>
</tr>
<tr>
<td>Age*CM</td>
<td>1</td>
<td>0.45600553</td>
<td>0.45600553</td>
<td>0.16</td>
<td>0.6932</td>
</tr>
<tr>
<td>Income*CM</td>
<td>1</td>
<td>4.40326359</td>
<td>4.40326359</td>
<td>1.50</td>
<td>0.2202</td>
</tr>
<tr>
<td>Age*Gender</td>
<td>1</td>
<td>1.46682070</td>
<td>1.46682070</td>
<td>0.50</td>
<td>0.4792</td>
</tr>
<tr>
<td>Income*Gender</td>
<td>1</td>
<td>1.39829186</td>
<td>1.39829186</td>
<td>0.48</td>
<td>0.4896</td>
</tr>
<tr>
<td>Age*Income</td>
<td>1</td>
<td>0.18635540</td>
<td>0.18635540</td>
<td>0.06</td>
<td>0.8009</td>
</tr>
<tr>
<td>Age<em>Income</em>CM*Gender</td>
<td>3</td>
<td>3.84045050</td>
<td>1.28015017</td>
<td>0.44</td>
<td>0.7265</td>
</tr>
</tbody>
</table>

All input variables and their interactions are not statistically significant to physician office visits.

For the variable Rx, which is the number of pharmacy prescription claims, the results are displayed in Table 4.13. The model is significant with a p-value <.0001.

Table 4.13. Regression model: Dependent variable Rx vs. input variables

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>13</td>
<td>1052.53</td>
<td>80.963</td>
<td>5.16</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9963</td>
<td>156284.44</td>
<td>15.686</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9976</td>
<td>157336.96</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>1</td>
<td>17.3910967</td>
<td>17.3910967</td>
<td>1.11</td>
<td>0.2924</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>3.1446016</td>
<td>3.1446016</td>
<td>0.20</td>
<td>0.6544</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>39.1684699</td>
<td>39.1684699</td>
<td>2.50</td>
<td>0.1141</td>
</tr>
<tr>
<td>Income</td>
<td>1</td>
<td>65.6563137</td>
<td>65.6563137</td>
<td>4.19</td>
<td>0.0408</td>
</tr>
<tr>
<td>CM*Gender</td>
<td>1</td>
<td>81.2390475</td>
<td>81.2390475</td>
<td>5.18</td>
<td>0.0229</td>
</tr>
<tr>
<td>Age*CM</td>
<td>1</td>
<td>28.2443578</td>
<td>28.2443578</td>
<td>1.80</td>
<td>0.1797</td>
</tr>
<tr>
<td>Income*CM</td>
<td>1</td>
<td>1.3040926</td>
<td>1.3040926</td>
<td>0.08</td>
<td>0.7731</td>
</tr>
<tr>
<td>Age*Gender</td>
<td>1</td>
<td>2.1682819</td>
<td>2.1682819</td>
<td>0.14</td>
<td>0.7101</td>
</tr>
<tr>
<td>Income*Gender</td>
<td>1</td>
<td>0.8473760</td>
<td>0.8473760</td>
<td>0.05</td>
<td>0.8162</td>
</tr>
<tr>
<td>Age*Income</td>
<td>1</td>
<td>39.6876279</td>
<td>39.6876279</td>
<td>2.53</td>
<td>0.1117</td>
</tr>
<tr>
<td>Age<em>Income</em>CM*Gender</td>
<td>3</td>
<td>284.4695263</td>
<td>94.8231754</td>
<td>6.04</td>
<td>0.0004</td>
</tr>
</tbody>
</table>
The input variables enrollment in CM, Gender and age are not statistically significant with a p-value higher than 0.05 level of significance. Income was significant at 0.05 significance level well as the interaction between CM and gender and the interaction between age, income, CM and gender.

For the ICU variable, number of ICU stays claims, the clean dataset had no ICU claims, so no results to be displayed for this variable.

Also, the selected study population didn’t have admission claims, there was no result for the dependent variable hospital admission.

For the urgent care clinics visits, the variable urgent vs other input variables model is displayed in Table 4.14. The model is significant with a low p-value that is <.0001.

Table 4.14. Regression model: Dependent variable urgent vs. input variables

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>13</td>
<td>0.618</td>
<td>0.047</td>
<td>3.17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9963</td>
<td>149.318</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9976</td>
<td>149.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>1</td>
<td>0.024</td>
<td>0.02451335</td>
<td>1.64</td>
<td>0.2010</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>0.001</td>
<td>0.00057399</td>
<td>0.04</td>
<td>0.8448</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>0.018</td>
<td>0.01804803</td>
<td>1.20</td>
<td>0.2725</td>
</tr>
<tr>
<td>Income</td>
<td>1</td>
<td>0.105</td>
<td>0.10484039</td>
<td>7.00</td>
<td>0.0082</td>
</tr>
<tr>
<td>CM*Gender</td>
<td>1</td>
<td>0.037</td>
<td>0.03707976</td>
<td>2.47</td>
<td>0.1158</td>
</tr>
<tr>
<td>Age*CM</td>
<td>1</td>
<td>0.003</td>
<td>0.00283862</td>
<td>0.19</td>
<td>0.6634</td>
</tr>
<tr>
<td>Income*CM</td>
<td>1</td>
<td>0.032</td>
<td>0.03234035</td>
<td>2.16</td>
<td>0.1419</td>
</tr>
<tr>
<td>Age*Gender</td>
<td>1</td>
<td>0.004</td>
<td>0.00369444</td>
<td>0.25</td>
<td>0.6196</td>
</tr>
<tr>
<td>Income*Gender</td>
<td>1</td>
<td>0.0001</td>
<td>0.00005297</td>
<td>0.00</td>
<td>0.9526</td>
</tr>
<tr>
<td>Age*Income</td>
<td>1</td>
<td>0.025</td>
<td>0.02460324</td>
<td>1.64</td>
<td>0.2001</td>
</tr>
<tr>
<td>Age<em>Income</em>CM*Gender</td>
<td>3</td>
<td>0.013</td>
<td>0.00449627</td>
<td>0.30</td>
<td>0.8254</td>
</tr>
</tbody>
</table>

Age, gender, enrollment in case management were not significant in the regression model as well as the interaction of all the variables. However, Income was the only significant input variable with a p-value of 0.0082 which is less than the 0.05 level of significance. This means that median income has a statistically significant effect on the number of urgent clinics claims. To further investigate how income affects the urgent clinics claims, chi-square tables were generated. First, the population was classified into four classes based on the four quartiles of population median income in the dataset, and the following hypothesis was tested:

$H_0$: Median income and number of urgent clinics visits are independent

$H_1$: Median income and number of urgent clinics visits are not independent
The p-value for median income vs number of urgent clinics visit is 1.51E-10 which is lower than 0.05 level of significance, so we reject the null hypothesis.

Table 4.15 displays the chi-square expected vs. actual values. The bolded values in the first row show the actual values of urgent clinics claims in the data and the second row show the expected values for each quartile.

**Table 4.15.** Chi Square results: How income is affecting the number of urgent clinics claims, actual values vs expected values

<table>
<thead>
<tr>
<th>Median income classification/Urgent claims</th>
<th>No Urgent</th>
<th>Urgent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_Poor</td>
<td>2485</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2472.2</td>
<td>21.7</td>
</tr>
<tr>
<td>2_Lower_Mid_Class</td>
<td>2475</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>2472.2</td>
<td>21.7</td>
</tr>
<tr>
<td>3_Upper_Mid_Class</td>
<td>2484</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2472.2</td>
<td>21.7</td>
</tr>
<tr>
<td>4_High</td>
<td>2446</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>2473.2</td>
<td>21.7</td>
</tr>
</tbody>
</table>

Theoretically people with high income are expected to make approximately 21 visits to the urgent care clinics, however, the actual data shows that patients with high income had 49 urgent clinic claims; this means that patients in our dataset with high income tend to utilize urgent care clinics more than expected. On the other hand, patients with lowest income quartile had only 9 urgent care clinics claims, this indicates that patients with lowest income level, tend to utilize urgent care clinics less than expected.

The chi-square result shows that there is a difference in the preference in visiting the urgent care clinic based on income level and patients with higher income are more likely to utilize urgent care clinics based on the presented chi-square tables.

### 4.5 Neural networks (NN) vs linear regression in prediction of total cost

As a side result we built a predicting model for the total cost, based on the other input variables using machine learning tools and specifically an Artificial Neural Network. The R statistical software was used to create the Neural Network and we compared its accuracy against a simple Linear Regression predictor.

Generally, it is expected that a larger number of layers and nodes in a Neural Network increases the prediction accuracy but at the cost of computational speed and memory resources. For this study and in an effort to maintain a balance between efficiency and accuracy, we chose an architecture of two hidden layers with 7 and 6 nodes each. The architecture was informed by an initial analysis of possible architectures on smaller datasets of 1000 points, where various
configurations were tried. Although 7, 6 was not always the best in terms of prediction, it always scored among the top 5 models.

In order to have a robust analysis of our predictive method, we repeated the process, 40, times using initially subsets of 5,000 data points.

The training set had 4,500 (90%) points and then it was tested on the remaining 500 points (10%). The neural network prediction was compared to that of a multilinear regression model on the same learning and test sets. The $l_2$ norm (distance) between the predicted value and the real value was used as an indicator of better prediction, for each of the subsets.

The average distance between predicted value by NN and true value is 1.3619 and the average distance between linear model predicted value and true value is 1.1803. Out of the 40 repetitions we tried, the NN outperformed Linear Regression 15 times. This implies that when using 5,000 datapoints, for the most part, the LR model is outperforming the NN model in predicting the total cost.

The same process was repeated for datasets of 6,000 data points. Again, we compared the accuracy of the two models in 40 repetitions. This time, the NN outperformed the linear regression 32 times out of 40. So, with 6,000 data points, the NN starts outperforming the regression model.

Finally, the process was repeated for subsets of 8,000 data points, again for 40 repetitions. This time the NN outperformed the regression model 35 times. The average distance of the predictions of the neural network from the true value were 1.033 and that of the linear model 1.317.

This implies that our NN outperforms the linear model in the prediction of the total cost based on all the other input variables as the number of data points increase. The proposed threshold point based on this dataset is somewhere close to 6,000 points. It is expected that NN’s will continue to outperform linear regression as the number of data points increases. The predictive model we have created can now be used to predict the Total Cost an individual will be facing based on their characteristics, with a high enough accuracy. Figure 4.5 is a visual representation of one of the networks we created.

Given a patients’ demographics (age and gender) and socio-economic indicator (zip code or median income) along with patient’s history in utilizing health services in terms of number of claims and billed amount of each claim, the NN model can attempt to predict the future total cost of the patient. This prediction can be then used to decide whether this patient should be enrolled in case management or not. Alternatively, a linear regression model can be used with similar input and output. The question on which model to use lies in the size the available dataset. For a small number of data points, a linear model will yield more accurate prediction but for larger data sets (around or exceeding 6,000 data points), NNs will yield more accurate predictions.
Figure 4.5. Visual representation for one of the neural networks
Chapter 5. Discussion and Limitations

First of all, we need to make it clear that the results of this study are applied to the provided dataset, and although the give us good indicators of the general population’s behavior we do not claim that our results are immediately generalizable.

This study found that asthma patients total cost is primarily driven by the billed amount for the following claims: primary physician office visits, emergency room visits, pharmacy prescriptions and urgent care clinic visits. Furthermore, the data analysis showed that there is a difference in total cost with respect to gender, and male patients have a higher asthma related cost compared to female patients. Since the majority of our study population is underaged children, this finding is consistent with another study that reported that younger male patients who are Medicaid insured have higher ER visits, which translates to cost (Nath & Hsia, 2015).

With respect to factors associated with frequent utilization of primary health services, we found that patient income and enrollment in case management were the significant factors in predicting the number of emergency room visits. Similar studies that looked at the effect of gender and age on emergency room visits, reported that emergency room visits were influenced by age, where younger patients had more visits and it decreased with age (Baibergenova et al., 2005).

For urgent care clinics visits, patient’s median income was a significant factor, patients who had higher income were more likely to utilize urgent care clinics. A previous study related socio-economic factor to another health resource utilization, and the study found that lower income is associated with higher risk of hospitalization (Eisner et al., 2001).

The comparison in total cost between patients enrolled in case management vs patients not enrolled showed that there was no real change over time in the mean of difference in cost upon enrollment in case management. This could indicate that asthma patients who were enrolled in case management do require to be case managed as they do have a higher overall cost. However, it also raises a question about the degree of effectiveness of the current approach of case management. This study proposed common success factors for case management as collected from literature.

For the total cost prediction model, the linear regression outperformed NN in predicting total cost for the 5,000 datapoints run, however, the NN stated outperforming linear regression as the number of datapoints increased, this study proposes the threshold of datapoints needed to train NNs to outperform a linear regression model in prediction at around 6,000 points. It is also expected that the larger number of data points, the better the NN will perform compared to linear regression.

There were several limitations of this study, first, patient factors such as marital status, race and ethnicity weren’t captured in the claims data. Secondly, patients asthma severity level was not indicated in the data which limited our ability to examine the effect of asthma severity on
cost. Third, our dataset was only for Medicaid insured patients in Louisiana, which limits the external validity of this study findings, our findings cannot be generalized and applied to other types of insurance such as private insured patients. And lastly, the study duration is limited, as case management service information wasn’t captured in the data system before 2015.
Chapter 6. Conclusion

This study conducted an analysis of Medicaid insured patients in Louisiana. It was found that patients with history of claims related to emergency room visits, pharmacy prescription, urgent care clinic visits and primary physician claims have a higher overall cost. The case management enrollment has no significant effect on the total cost over time for patients who are case managed compared to those that are not. Median income was significant in predicting the number of emergency room, urgent care clinics and pharmacy prescription claims. And finally, as side result, this study built two prediction models for total cost, for low number of data points, linear regression model was able to predict total cost with a good enough accuracy and the threshold number of points in which the neural network outperforms linear regression is about 6,000 point, so for higher number of data points, neural networks was a better prediction model, and the accuracy of neural network prediction is expected to increase as the number of data points increases.

This study is a foundation for further future analysis of asthma patients demographics, claims history and their association with health resources utilization. Excessive health resources utilization is an indication of poor disease management and places a burden on the economy and society. Future work is to collect patients information and diversify the source of data to include multiple insurance types from different geographical regions and patients from all age groups. Furthermore, future analysis should be conducted over longer period of time to better examine the effectiveness of case management in reducing overall cost. Finally, future studies can also look at other demographic and clinical factors that weren’t captured in this study such as race, education level, marital status and asthma severity level to examine their effect on asthma overall cost and level of utilization of primary health services.
Appendix. IRB Approval

ACTION ON PROTOCOL APPROVAL REQUEST

TO: Isabelina Nahmens  
    Mechanical and Industrial Engineering

FROM: Dennis Landon  
      Chair, Institutional Review Board

DATE: November 2, 2017

RE: IRB# 3682

TITLE: Identifying Key Factors Associated with High Risk Asthma Patients to Reduce Cost of Health Resources Utilization


Review type: Full ___ Expedited _X___  
Review date: 11/1/2017

Risk Factor: Minimal ___ X ___ Uncertain ___ Greater Than Minimal ___

Approved ___ X ___ Disapproved ___

Approval Date: 11/2/2017  Approval Expiration Date: 11/1/2018

Re-review frequency: (annual unless otherwise stated)

Number of subjects approved: N/A

LSU Proposal Number (if applicable):

Protocol Matches Scope of Work in Grant proposal: (if applicable)

By: Dennis Landon, Chairman  

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING — Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU’s Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report) prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
8. SPECIAL NOTE: When emailing more than one recipient, make sure you use bcc.

*All investigators and support staff have access to copies of the Belmont Report, LSU’s Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb


References


Centers for Disease Control and Prevention. (2010). *National Hospital Ambulatory Medical Care Survey: Outpatient Department Summary Tables* (National Hospital Ambulatory Medical Care Survey). Retrieved from https://www.cdc.gov/asthma/most_recent_data.htm


Vita

Amani Ahmad, completed her bachelor’s degree in Industrial Engineering from the University of Jordan. She worked for one year as a Supply Chain Management Coordinator in Samsung at Amman, Jordan. As her career started to progress, she planned to learn more about data analytics and decided to move to the United States to pursue graduate studies. Upon completion of her master’s degree, she will begin working in IT consulting.