PREDICTING FREQUENT EMERGENCY DEPARTMENT VISITS AMONG CHILDREN WITH ASTHMA USING ELECTRONIC HEALTH RECORD DATA

A Thesis

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Lala Tanmoy Das

August 2015

ABSTRACT

Background: Asthma is one of the most common chronic conditions among children and is the third leading cause of pediatric hospitalization among children under age 15. Asthma-related emergency department (ED) visits are common and expensive to the health system. One of the proposed cost savings measures is for hospitals to hire care managers who can coordinate care for individuals with chronic diseases such as asthma, for example by educating families about general preventative care practices, reconciling medications, and answering questions. It would be valuable to prospectively identify children likely to be frequent ED users, in order to enroll in care management programs. It is unclear if electronic health record (EHR) data can be used to predict which patients will frequently use the ED for asthma.

Objective: To explore the predictability of frequent ED use among pediatric asthma patients in New York City using data from an EHR from one medical center.

Methods: We performed a literature review and interviewed 5 physicians affiliated with Weill Cornell Medical Center to generate a list of potential predictors of ED use. We operationalized a subset of these predictors from an EHR system. We performed bivariate statistics to examine the unadjusted relationship between each variable and frequent ED use. Then we evaluated and compared the performance of several machine-learning algorithms to predict which children with asthma will use the ED two or more times in the next year. The algorithms we used were: logistic regression best subsets, Least Absolute Shrinkage and Selection Operator (LASSO) regression, Random Forests, Classification and Regression Trees (CART), and Support Vector Machines (SVM). We evaluated model performance based on Area Under the Curve

(AUC), positive predictive value (PPV), sensitivity, calibration, and classification error.

Results: We operationalized 52 predictors. Bivariate analysis showed significant associations between many of the clinical and demographic predictors and frequent ED use. All the predictive algorithms performed similarly, with very good area under the curve (AUC) values, but poor positive predictive value and sensitivity. We selected a two variable model as our final model based on the predictors that appeared significant across the algorithms: number of ED visits in the previous year and type of insurance. Publicly insured patients with asthma who used the ED four or more times in the baseline year have a 50% or greater probability of being a frequent ED user in the following year. The same utilization pattern is seen among privately insured patients who have six or more ED visits in the baseline year.

Conclusions: Children who are currently frequent users of the ED are likely to continue to do so. The threshold for identifying these children is lower among children with public insurance compared to those with private insurance. A two variable (prior ED visits and insurance status) model to predict which children with asthma will be future ED users is as accurate as predictions from several machine learning algorithms. These observations can be used to identify children with asthma who may benefit from enrollment in a care management program, using data from an EHR.

BIOGRAPHICAL SKETCH

Lala Tanmoy Das graduated from DePauw University in 2012 with a Bachelor of Arts degree in Biological Sciences. He enrolled in Weill Cornell's Master of Science in Health Informatics program in 2014.

ACKNOWLEDGEMENTS

I would like to thank several people who have provided support, guidance, and encouragement during the entire course of this research project and the writing of this thesis.

I would like to thank my adviser Zachary Grinspan for being an incredible mentor, supervisor, and editor. His words of advice and suggestions have been extremely valuable at every step along the way, and I have grown to be a better researcher because of him.

I would like to thank Erika Abramson for all the brainstorming ideas, thesis comments, and edits.

I would like to thank my sister Meenakshi Das Lala for taking care of me through the many late nights of coming home after endless amounts of data analysis. Thank you to Alicia Suarez for not only introducing me to health policy research but also igniting a passion for it. Shout out to Brandy Richmond and Aruni Roy Chowdhury for being a huge source of support. Thank you to David Christini, Trine Krogh-Madsen, Tashalee Brown, Bonnie Quach, Francis Ortega, my friends, and my wonderful parents for cheering me on!

Finally, I am grateful to my partner Eric Kutscher for his love, support, kindness, and words of encouragement at all hours of the day during graduate school and this thesis process. I can never thank you enough for being by my side and always bringing a smile on my face especially when things got stressful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

Asthma is one of the most common chronic conditions among children and is the third leading cause of pediatric hospitalization among patients under the age of 15.[1] More than 7.1 million (over 10%) children living in the United States had asthma in 2011.[2] Half of children with asthma have at least one exacerbation per year, which often necessitate ED visits and inpatient hospital admissions.[3] In 2010 alone, there were 640,000 asthma-related pediatric ED visits.[1]

Pediatric asthma care is expensive to the US healthcare system. In 2007, the U.S. spent over $56 billion on asthma care, half of this on pediatric asthma alone.[4] State Medicaid programs combined spent in excess of $272 million on pediatric asthma-related ED visits in the United States in 2010.[5] Importantly, patients with poorly controlled severe asthma cost nearly $5000 more per patient per year compared to average pediatric asthma costs.[6]

Asthma is an ambulatory care sensitive condition,[7] which implies that timely and effective ambulatory care can reduce emergency department (ED) visits and hospitalizations. High quality outpatient care might include providing education to patients and their family members, controlling disease severity with daily medication, helping families make lifestyle adjustments to avoid triggers, and managing acute exacerbations in the home or in an office setting. Frequent ED visits for asthma and

high rates of inpatient hospitalization in a community may indicate insufficient preventative measures, shortage of primary care access, or poor performance of care delivery systems. One strategy that has the potential to improve disease control is the implementation of targeted care management (CM).

In order to reduce preventable ED use, many healthcare organizations have created CM teams.[8] Care managers work directly with patients to coordinate care services associated with hospital discharge, provide information about their disease and preventive care measures, reconcile medications, answer post-hospitalization follow-up questions, and manage treatment-related services with other healthcare providers.[8] Well-implemented care management programs can reduce hospital use, lower costs, and improve clinical outcomes for patients with chronic diseases.[9,10] Despite these positive outcomes, identification of the best patient cohort for potential enrollment in CM programs remains a challenge due to insufficient published research and a lack of consensus among administrators, healthcare providers, and policy experts to guide enrollment strategy.

Given the challenges associated with patient cohort identification, one proposed way to better identify patients at-risk for frequent ED use is to use predictive modeling techniques.[11] A predictive model takes quantifiable data elements from a patient's medical history and makes a prediction about future healthcare use, such as ED use. Predictive models may use classical methods like regression or more sophisticated machine learning techniques to predict outcomes. Predictive models can

generate a risk score to identify patients likely to make frequent ED visits based their medical history. Subsequently, a risk score threshold can be used to determine a patient's inclusion or exclusion into CM programs. Frequent ED use can also be predicted in a binary sense; that is, the model will predict the outcome as "yes" or "no". Patients who are predicted to be frequent ED users may be enrolled in CM.

Investigators have used PM to identify several factors among children with asthma that are associated with frequent ED use. These factors include prior healthcare utilization history,[12] demographics,[13] co-morbidities,[14] insurance status,[15] and medication history.[16] For example, frequency of asthma-related ED visit in the previous year was associated with the frequency of asthma-related ED visit in the following year.[12] Another study showed that controller-to-asthma medication ratio significantly predicted frequency of future ED visit, and hospitalization.[16] A recent study also found that living in a single-parent home, having a disability related to asthma, and asthma severity are significant predictors of frequent ED visits.[13] However, a 2011 review found that many risk prediction models have only modest levels of accuracy.[17]

In the past, most investigators have used regression as a predictive modeling tool. While there are several benefits of using regression, there are important limitations. For example, regression techniques have difficulties accounting for missing data, large numbers of variables, non-linear relationships between variables, and unusual distributions – all commonly seen in healthcare data. Since then,

significant efforts have been made to optimize predictive performance by using newer techniques that can better handle the limitations associated with regression. The results have shown slight improvements in discriminative accuracies,[18-21] warranting more research.

Modeling outcomes, however, depend not only on analytic methods but also on the quality of the data source and the extent of accessible data elements. Many studies use administrative data from statewide or national databases.[22] While there are some advantages to this approach such as ease of availability, greater number of subjects, and easier assessment of medication compliance, there is very limited information available about clinical factors pertaining to patients in administrative datasets. Clinical factors may be crucial to better understanding patient's health utilization patterns. Thus, an alternative approach is to use data from electronic health records (EHRs). In addition to many of the data elements available in administrative datasets, such as visit history and diagnosis codes, EHRs contain detailed clinical information such as lab test results, medical comorbidities, and disease severity.[23]

In this study, we explore the predictability of frequent ED use among pediatric patients in New York City using data from an EHR from one medical center. Our goals were twofold. First, we created and validated a database of factors derived from an EHR system that are known or presumed to predict frequent ED use among pediatric asthma patients. Our second goal was to evaluate and compare the

performance of several machine-learning algorithms to predict which children with asthma will use the ED four or more times in the next year.

CHAPTER TWO
METHODS

Study Design Overview. We performed a retrospective cohort study in order to investigate factors associated with frequent ED use. Patients' healthcare utilization data for predictive modeling were collected for years 2013 (Year 1) and 2014 (Year 2) using the outpatient EHR of one medical center in New York City. The Weill Cornell Medical Center Institutional Review Board approved the study.

Potential Predictors. We performed a literature search to identify reported predictors of frequent ED use among pediatric asthma patients. In addition, in order to make a more comprehensive list of predictors for predictive modeling, we interviewed a convenience sample of physicians who were affiliated with Weill Cornell Medical College and had clinical experience working with pediatric asthma patients.
The interviews were semi-structured. Examples of questions we asked all physicians were: (1) What do you think are the common reasons why children with asthma frequently use the ED? (2) What socio-demographic factors contribute to frequent ED use? (3) What clinical markers, in your opinion, may correlate with frequent ED use among children with asthma? Specific follow-up questions were asked based on their responses. We took hand-written notes during the interviews that were later used to compile and generate a list of potential predictors.

Inclusion Criteria. We included all individuals below 18 years of age who were patients at Weill Cornell Medical Center's outpatient clinic, prescribed Albuterol or Xopenex, or with an ICD-9 code of 493.x (asthma) in 2013.[24]

Predictor Variables. Data for each patient included demographic variables such as age, sex, race and ethnicity, borough of New York, type of insurance, access to primary care physician, prior health service utilization history like number of ED, medical complexity using the Pediatric Medical Complexity Algorithm (PMCA),[25] inpatient, and outpatient visits, clinical factors such as Body Mass Index (BMI), IgE blood test, dependence on technology for asthma care, and allergies, variables for psychiatric and medical comorbidities, and medication usage.

Outcome. We define frequent ED use as ≥ 2 ED visits. This is less than the more common definition of 4+ visits,[26-28] which was a very rare event in our data (1.5% of patients).

Bivariate Analysis. We investigated associations between predictor variables and the frequent (2+) ED visits in Year 2. We performed Pearson's correlation test for continuous predictor variables. For categorical predictors, we used the Chi-squared test.

Predictive Modeling Strategy. We used two modified regression techniques, both of which provide easily interpretable results. We also used two decision tree algorithms

and support vector machines which are more challenging to interpret, but have the potential to provide more accurate predictions.

We randomly split the data in half, creating a training and test set. We applied the following techniques to the training set:

Logistic Regression, Best Subsets. We created 1,2, and 3-variable models using the best subsets algorithm and logistic regression. This technique identifies the best-fitting regression model by examining all combinations of predictor variables that are specified. This technique is computationally feasible for large numbers of variables when the total the number of allowed predictor variables is restricted to a smaller number.[29]

Regularized Logistic Regression, Lasso. The LASSO (Least Absolute Shrinkage and Selection Operator) is a regression method that involves penalizing the absolute size of regression coefficients, as they grow too large. By constraining the sum of the absolute values of the estimates, some of the parameter estimates become zero, thereby limiting the number of variables in the final model. The LASSO requires a parameter λ, which we determined via cross-validation within the training set. [29,30]

Decision Trees. We used two decision tree algorithms. CART (Classification and Regression Tree) algorithm constructs and prunes a decision tree

automatically.[31] Since our outcome variable is binary, we used the classification method in our algorithm. Random Forests builds a series of trees using random subsets of the training data.[32] When a new example is introduced, each tree of the Random Forest "votes" on the outcome. The outcome with the majority vote is the final prediction.

Geometric. Support vector machines is a supervised learning model based on geometric principles. The algorithm represents every example, belonging to one of two categories, as a point in space such that examples of either category are separated by a clear gap that is as wide as possible. When a new example is introduced the predicted outcome is determined based on which side of the gap it appears.[33]

Evaluation of Models. After building the models with the training set data, we evaluated models chosen by each technique on the test set using the following criteria: (1) area under the receiver operating curve (AUC), (2) calibration error, (3) classification error, (4) sensitivity, and (5) positive predictive value (PPV). Calibration error is defined as the measurement of mean absolute error between predicted probability and observed proportion calculated as follows:

a) Split individuals into 5 groups, based on evenly distributed bins of predicted probability (i.e. 0-20%, >20-40%, etc.)
b) Calculate the mean absolute difference between the predicted probability and the observed proportion of frequent ED use, across these 5 groups.

Classification error was defined as (1-Accuracy) * 100, where Accuracy is defined as percentage of times the predicted outcome is equal to the observed outcome.

The following rubric was used to evaluate each technique's model AUC, sensitivity, and PPV: <0.6 poor, 0.6-0.69 fair, 0.7-0.79 good, 0.8-0.89 very good, and 0.9-1.0 excellent. For calibration and classification error evaluation, we use the following rubric: 0-4% excellent, 5-9% very good, 10-14% good, 15-20% fair, and >20% poor.

Model Selection. We selected a final model based on sparsity (i.e. small number of variables), accuracy, and interpretability.

Additional Analysis. Based on the variables in the final model, we performed an additional analysis to investigate the relationship between ED visits in the baseline year and ED visits in the follow-up year among patients with public insurance and private insurance. We determined the number of ED visits in Year 1 that corresponded to a 50% or greater chance of frequent ED use in Year 2, in order to suggest a cutoff for a parsimonious prediction rule.

Statistical Software. We used the R software package (version 3.1.1)[34] for analysis, using the following packages: "gdata",[35] "randomForest",[32] "rpart",[31] "rpart.plot",[36] "glmnet",[30] "ROCR",[37] "caret",[38] "e1071",[33] "ggplot2",[39] "data.table",[40] "leaps",[41] "stats",[34] "pROC".[42]

CHAPTER THREE
RESULTS

List of Potential Predictor Variables. We reviewed 17 recent papers[1-5, 6-8, 12-16, 25, 43-45] and identified 4 categories of factors that are associated with frequent ED use: (1) demographics; (2) prior healthcare utilization history; (3) co-morbidities; and (4) medication history. The demographic variables were race, ethnicity, insurance status, and having access to a primary care physician. Healthcare utilization history variables were (1) total number of ED visits and (2) total number of outpatient visits. The co-morbidities variables were: (1) food and other allergies; (2) congenital lung abnormalities, and (3) gastro-esophageal reflux disease (GERD). For medication history we had the variable Controller-to-asthma medication ratio. We interviewed five physicians: three general pediatricians, one pediatric pulmonologist, and one pediatric emergency medicine physician. We compiled a list of 74 potential predictors from the literature review and the interviews. (Table 1)

Operationalization of Variables in EHR. We were able to operationalize 36 out of 74 predictors from Weill Cornell Physician Organization's outpatient electronic health record (EHR) data. Based on data available in the EHR, we added 16 additional asthma-related variables such as medical and psychiatric comorbidities, specific allergies, and use of asthma medication delivery devices. The final number of predictors was 52.

Table 1. List of variables from literature review, physician interviews, and operationalized predictors.

	Proposed Variable	Operationalized Variable
DEMOGRAPHICS	Individual	
	Age	Age
	Sex	Sex
	Race*	Race
	Ethnicity*	Ethnicity
	Insurance Status*	Insurance Type
	Socioeconomic Status	Borough of New York
	Frequent change of address	-
	Yes/No for Homelessness/Living in a shelter	-
	Yes/No and Count to Recommend to Social Worker	-
	Family Characteristics	
	Single parent family	-
	Degree of parental health education	-
	Family history of allergic disease	-
	Zip Code Characteristics	
	Population density of zip code	-
	Degree of education	-
	Zip code	Borough of New York
	Environmental Factors	
	Yes/No for Exposure to secondhand smoke	-
	Yes/No for parents/guardians who smoke	-
	Yes/No to dust exposure	-
	Allergens/Trigger exposure at home	Allergy List (Categories)
CLINICAL DATA	Disease Factors	

	IgE levels	IgE Test
	Asthma Severity	-
	Abnormal PFTs (Pulmonary Function Tests)	-
	FEV1 (Forced Expiratory Volume in 1 second)	-
	BMI	BMI
	Respiratory Viral Panel	-
	Allergy Test	Allergy List (Categories)
	Yes/No to Flu Shot past year	-
MEDICATIONS	Medication refill history	-
	Controller-to-Asthma medication ratio*	-
	Yes/No for Inpatient steroids	-
	Yes/No for outpatient steroids	-
	Xolair	-
	-	Yes/No for Asthma Medication Device
	Advair	Yes/No for Inhaled Drug Combination
	Symbicort	Yes/No for Inhaled Drug Combination
	Albuterol use	Yes/No for Inhaled Rescue Medication
	Yes/No for prescription beta agonists	Yes/No for Inhaled Rescue, Yes/No for Inhaled Drug Combination
	Yes/No for inhaled steroids	Yes/No for Inhaled Steroids
	Flovent (Fluticasone)	Yes/No for Inhaled Steroids
	Qvar (Beclomethasone)	Yes/No for Inhaled Steroids
	Pulmicort (Budesonide)	Yes/No for Inhaled Steroids
	Yes/No for leukotriene inhibitors	Yes/No for Leukotrienes
	Montelukast (Singulair)	Yes/No for Leukotrienes
	-	Yes/No for Mast Cell Stabilizer
	Yes/No for oral steroids	Yes/No for Oral Steroids

	Prednisolone	Yes/No for Oral Steroids
	Prednisone	Yes/No for Oral Steroids
	-	Yes/No for Other Medications
COMORBIDITIES	PMCA	PMCA (Stringent) Classification
	-	Yes/No for Allergic Rhinitis
	Food and other allergies*	Yes/No for Allergies
	-	Yes/No for Anxiety
	-	Yes/No for Bronchiolitis
	-	Yes/No for Bronchitis
	Chronic Lung Disease	Yes/No for Bronchopulmonary Dysplasia
	Broncho-pulmonary Dysplasia	Yes/No for Bronchopulmonary Dysplasia
	Congenital lung abnormalities*	Yes/No for Bronchopulmonary Dysplasia
	-	Yes/No for Depression
	Eczema	Yes/No for Eczema
	-	Yes/No for Gastrostomy Tube
	Gastroesophageal Reflux Disease (GERD)*	Yes/No for GERD
	-	Yes/No for Pneumonia
	Premature birth	Yes/No for Premature Birth
	Tracheostomy Dependent	Yes/No for Tracheostomy
	-	Yes/No for Vocal Cord Dysfunction
PRIOR HEALTH SERVICE UTILIZATION	<u>Ambulatory & ED</u>	
	Total number of ED Visits*	2013 (Year 1) and 2014 (Year 2) ED visits
	Total number of ED Visits for Asthma	-
	Outpatient Visits, Total*	Year 1 and 2 Outpatient

	Visits
Outpatient Visits, General Pediatrics	-
Outpatient Visits, Pediatric Pulmonology/Allergist	-
Missed Outpatient Visits for Primary Care & Specialist	-
Yes/No for Having PCP*	Yes/No for Has PCP
Ratio of ED Visit/Primary Care Visit	-
Reliance on ED for Asthma Care	-
Time of ED Visit	-
Weekend/Weekday ED Visit	-
Inpatient	
Total Inpatient Admissions	Year 1 and 2 Inpatient Visits
Admission to ICU for Asthma	-
Admission to ICU, Total Number	-
Intubation for Asthma	-
Inpatient Admissions from ED	-
Number of Inpatient days	-
Number of ICU days	-
O2 levels during Inpatient stay	-
Maximal level of respiratory support required	-

*** Indicates variables found in literature review.**

Description of Cohort. There were 2691 patients in the final dataset. Median age was 6 [IQI: 3 – 11]. Sixty percent of patients were male, and 40% were female. Racially, 36% were White, 13% were Black or African American, 6% were Asian, and 1% American Indian or Alaska National. A large group (29%) fell into the category Other

Combinations not Described. Almost 22% identified as having Hispanic or Latin or Spanish origin. For borough of residence, 46% lived in Manhattan, 14% lived in Queens, 13% in Brooklyn, 12% in the Bronx, and 1% in Staten Island. The remaining 14% lived outside New York City. For insurance type, 67% of patients had private insurance, 28% were publicly insured, and the remaining 5% were unknown. A majority of patients had a primary care physician (78%). However, for almost 21% of the patients whether they had access to a primary care doctor was unknown.

Prevalence of Frequent ED users. There were 184 (6.8%) frequent ED users (≥ 2 visits) during Year 2 among the 2691 individuals in the full dataset. There were 98 (7.3%) frequent ED users in the training data, and 86 (6.3%) frequent ED users in the test data.

Bivariate Associations between Predictors and Frequent ED use. Several predictor variables were significantly associated with frequent ED use in the follow-up year. (Table 2) Females were more likely than males to be frequent ED users. Patients who racially identify as "Black or African American" or "Other" were more than 50% more likely to use the ED frequently. Other factors that contribute to a high ED use are: having Hispanic/Latino/Spanish origin, living in the Bronx or Queens, having public insurance, being a medically complex patient, being obese, and high health utilization in the baseline year. Among medical and psychiatric comorbidities, patients with allergic rhinitis, eczema, and pneumonia were more likely to be frequent ED users. Additionally, patients who have had gastrostomy tubes or tracheostomy for care

were at an elevated risk for ED use. Interestingly, frequent ED users were twice as likely to be allergic to cockroaches, but less likely to have mold or tree allergies. For medication usage, frequent ED users used asthma medication devices such as spacers and nebulizers almost 25% more than non-frequent ED users, took more inhaled rescue drugs, mast cell stabilizers, and other medications. (Table 2)

Table 2. Descriptive statistics of variables classified on the basis of frequent ED use.

	Variable	**< 2 visits (N = 2507)**	**> = 2 visits (N = 184)**	**p- value**
Age	Age	6 [3-11]	6 [3-11]	ns §
Sex	Female	996 (40%)	87 (47%)	0.05 ¶
	Male	1511 (60%)	97 (53%)	
Race	American Indian/Alaskan	0 (0%)	5 (3%)	< 0.001 ¶
	Asian	151 (6%)	3 (2%)	
	Black or African American	323 (13%)	36 (20%)	
	White	940 (37%)	30 (16%)	
	Other	696 (28%)	93 (50%)	
	Declined	354 (14%)	17 (9%)	
	Unknown	43 (2%)	0 (0%)	
Ethnicity	Hispanic/Latino/ Spanish Origin	520 (21%)	69 (38%)	< 0.001 ¶
	Not Hispanic/Latino/ Spanish	1461 (58%)	80 (43%)	
	Declined	421 (17%)	19 (10%)	
	Unknown	105 (4%)	16 (9%)	
Borough of New York	Bronx	303 (12%)	33 (18%)	< 0.001 ¶
	Brooklyn	327 (13%)	19 (10%)	
	Manhattan	1149 (46%)	76 (41%)	
	Queens	336 (13%)	48 (26%)	

	Staten Island	24 (1%)	0 (0%)	
	Other	368 (15%)	8 (5%)	
Insurance Type	Private	1749 (70%)	59 (32%)	< 0.001 ¶
	Public	635 (25%)	122 (66%)	
	Unknown	123 (5%)	3 (2%)	
Has Primary Care Physician (PCP)	Does not have PCP	16 (1%)	0 (0%)	< 0.01 ¶
	Has PCP	1980 (79%)	130 (71%)	
	Unknown	511 (20%)	54 (29%)	
Utilization Baseline Year (Year 1)	ED Visits	0 [0 - 0]	1 [0.75 - 3]	< 0.001 §
	Inpatient Visits	0 [0 - 0]	0 [0 - 1]	< 0.001 §
	Outpatient Visits	3 [1 - 6]	6 [3 - 11]	< 0.001 §
	Other Visits	0 [0 - 0]	0 [0 - 0]	ns §
PMCA (More stringent)	Chronic Disease, Medically Complex	277 (11%)	42 (23%)	< 0.001 ¶
	Chronic Disease	795 (32%)	72 (39%)	
	No Chronic Disease	1306 (52%)	67 (36%)	
	Other	129 (5%)	3 (2%)	
BMI	Normal	2080 (83%)	158 (86%)	< 0.01 ¶
	Overweight	123 (5%)	8 (5%)	
	Obese	72 (3%)	11 (6%)	
	Unknown	232 (9%)	7 (3%)	
IgE Test	Abnormal	54 (2%)	6 (3%)	ns ¶
	Normal	47 (2%)	3 (2%)	
	Not Sent	2406 (96%)	175 (95%)	
Psychiatric Comorbidities	Anxiety	1 (0.03%)	1 (0.5%)	ns §
	Depression	12 (0.5%)	2 (1%)	ns §
Medical Comorbidities †	Allergic Rhinitis	415 (16%)	39 (21%)	< 0.01 §

	Bronchiolitis	213 (8.5%)	15 (8%)	ns §
	Bronchitis	17 (1%)	0 (0%)	ns §
	Broncho-pulmonary Dysplasia	69 (3%)	7 (4%)	ns §
	Eczema	156 (6%)	18 (10%)	< 0.01 §
	Pneumonia	132 (5%)	17 (9%)	< 0.05 §
	Premature Birth	35 (1%)	3 (2%)	ns §
	Vocal Cord Dysfunction	11 (0.4%)	3 (2%)	ns §
Technology Dependence	Gastrostomy Tube	40 (1.5%)	9 (5%)	< 0.001 §
	Tracheostomy	38 (1.5%)	7 (4%)	< 0.01 §
Allergies	Animal	253 (12%)	16 (9%)	ns §
	Chemical	19 (0.8%)	3 (2%)	ns §
	Cockroach	47 (2%)	8 (4%)	< 0.01 §
	Drug	294 (12%)	27 (15%)	ns §
	Dust	140 (6%)	14 (8%)	ns §
	Food	506 (20%)	42 (23%)	ns §
	Fungus	5 (0.2%)	0 (0%)	ns §
	Herbal	12 (0.5%)	0 (0%)	ns §
	Latex	13 (0.5%)	2 (1%)	ns §
	Mold	1574 (62%)	102 (55%)	< 0.05 §
	Other	111 (4.4%)	9 (5%)	ns §
	Pollen	115 (5%)	11 (6%)	ns §
	Seasonal	20 (0.7%)	2 (1%)	ns §
	Trees	1663 (66%)	109 (59%)	< 0.05 §
Medication Usage *	Asthma Medication Device	565 (23%)	54 (29%)	< 0.05 §
	Inhaled Drug Combination	76 (3%)	8 (4%)	ns §
	Inhaled Rescue	2274 (91%)	177 (96%)	< 0.05 §
	Inhaled Steroids	1037 (41%)	82 (45%)	ns §
	Leukotrienes	294 (12%)	22 (12%)	ns §
	Mast Cell Stabilizer	4 (0.2%)	2 (1%)	< 0.05 §
	Oral Steroids	528 (21%)	43 (23%)	ns §
	Other Medications	1848 (74%)	157 (85%)	< 0.001 §

All values are either Median [Interquartile Range] or N (% of patients within the category of ED use in Year 2).
§ Pearson correlation. ED visits in Year 2 treated as a continuous variable.
¶ Chi-square test. ED visits in year 2 treated as 2-valued categorical variable (>= 2 or < 2 ED visits in Year 2).
ns Not significant at p-values < 0.05
† GERD and IgA were available to the predictive models but were 0% prevalent in both groups of ED use.

*** Medication Usage Classification:**
Asthma Medication Device = Aerochamber, Nebulizer, Nessi Spacer
Inhaled Drug Combination = Advair Diskus, Advair HFA, Budesonide-Formoterol, Combivent, Fluticasone-Salmeterol, Ipratropium-Albuterol, Symbicort
Inhaled Rescue = Albuterol, Ipratropium bromide, Ventolin, Xopenex
Inhaled Steroids = Beclomethasone, Budesonide, Flovent, Fluticasone, Pulmicort, Pulmoneb, Qvar
Leukotriene = Montelukast, Singulair
Mast Cell Stabilizer = Cromolyn
Oral Steroids = Orapred, Prednisolone, Prednisone, Dexamethasone
Other Medications = Non-asthma prescription and non-prescription medications

Predictability of Frequent ED Use. Qualitatively, all seven algorithms performed similarly. The area under the receiver operating curve (AUC) ranged from 0.81-0.87 for all algorithms excluding CART, indicating very good predictability. The performance of CART was fair with AUC of 0.66. The positive predictive value was mostly poor (< 60%), with the exception of LASSO, which had good performance (70%). The sensitivity was uniformly poor (16-27%). Calibration was good to fair (13-20%), with the exception of Random Forest, which had excellent calibration (4.1%). Classification error ranged from 5.8-6.6%. (Table 3)

The LASSO Model. The two variables selected by the LASSO technique (Table 3) were (1) number of ED visits in the baseline year, and (2) having public insurance. In

comparison to other models, it had the highest AUC (0.87), highest positive predictive value (70%) and lowest classification error (5.8%). Similar to other models it had poor sensitivity (16%) and calibration (20%).

Two Variable Model. The two variable logistic regression model based on the best subsets algorithm yielded the same variables in its final model as the LASSO. Compared to the LASSO, it had similar AUC (0.86) and classification error (5.9), better sensitivity (23%), and better calibration (13%). However, it had poorer positive predictive value (56%). The final variables in the pruned tree of the CART model were the same ones as the LASSO and two-variable best subsets. The three-variable best subsets model also had the two aforementioned variables, along with a third one (i.e. use of mast cell stabilizer).

Selection of Final Model. We selected a final model that had two predictors: (1) Number of ED visits in Year 1; and (2) Type of Insurance. Both of these variables emerged significant in the final parsimonious models of all the predictive algorithms.

Table 3. Evaluation of Predictive Modeling Techniques on a test set of 1346 patients.

Technique	Number of Variables	AUC	Calibration (%)	Classification Error (%)	Sensitivity (%)	PPV (%)	Variable Names
Logistic Regression, Best Subsets	1	0.84	13	5.9	27	59	Number of ED visits in Year 1
	2	0.86	13	5.9	23	56	Number of ED visits in Year 1, Insurance Type = Public
	3	0.86	14	6.1	23	56	Number of ED visits in Year 1, Insurance Type = Public, Use of Mast Cell Stabilizer = Yes
Lasso	2	0.87	20	5.8	16	70	Number of ED visits in Year 1, Insurance Type = Public
Random Forest	all (52)	0.83	4.1	6.4	19	52	-
CART	2	0.66	15	6.6	24	48	Number of ED visits in Year 1, Insurance Type = Public
Support Vector Machines	all (52)	0.81	14	6.0	19	62	-

The final plots illustrate frequent ED utilization in the follow-up year compared to the number of ED visits in the baseline year, separately among individuals with private and public insurance. Among the publicly insured, any individual with 4 or more ED visits in the baseline year had a 50% or higher probability of frequent ED use in Year 2 (Figure 1). Individuals, who visited the ED 4 or more times in the baseline year, used the ED, on average, 4 times in the following year. However, among the privately insured, 50% or higher probability of frequent ED use in Year 2 was attained at 6 or more ED visits during the baseline year (Figure 2). Individuals, who visited the ED 6 or more times in the baseline year, used the ED, on average, 11 times in the following year.

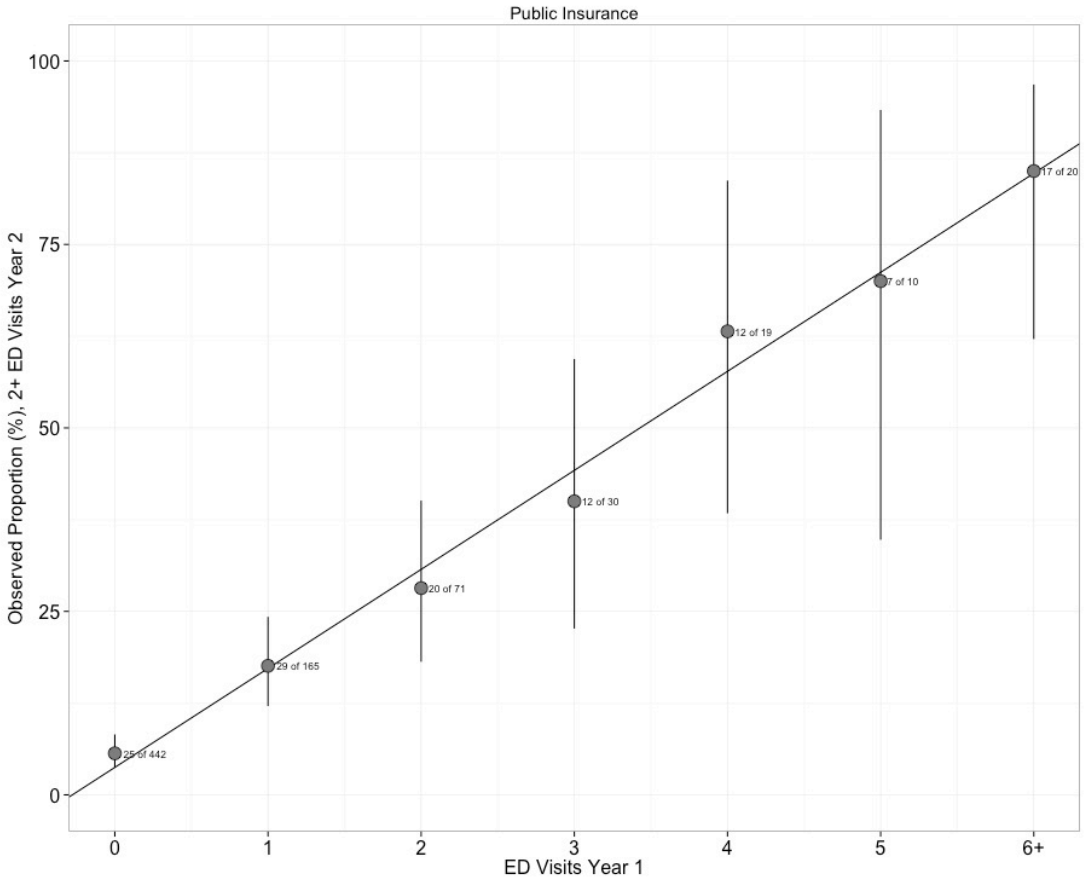

Figure 1. Relationship between number of ED visits in year 1 and observed proportion of frequent ED use in year 2 among publicly insured patients. The X axis denotes the number of ED visits in baseline Year 1. The Y axis denotes the corresponding proportion of patients (%) who were frequent ED users in the following year. Points indicate observed proportion of frequent ED users. The diagonal line shows the slope of the line using linear modeling. The error bars are the 95% confidence interval of the binomial distribution of the observed proportion.

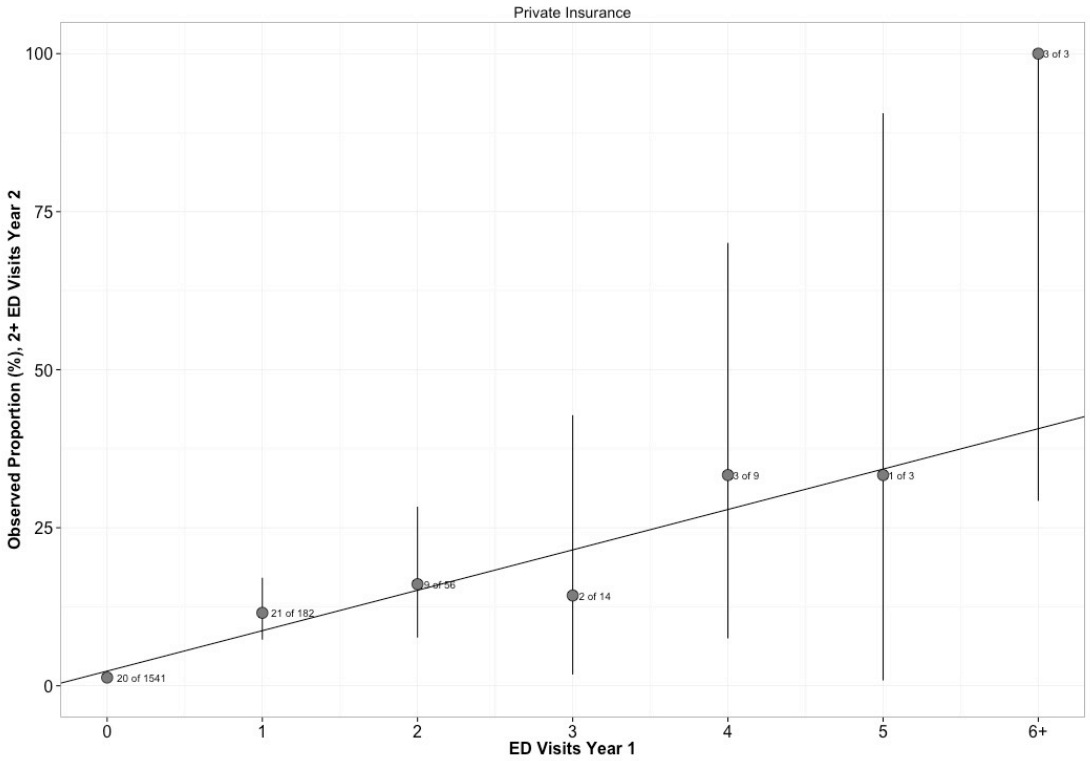

Figure 2. Relationship between number of ED visits in year 1 and observed proportion of frequent ED use in year 2 among privately insured patients. Points indicate observed proportion of frequent ED users. The X axis denotes the number of ED visits in baseline Year 1. The Y axis denotes the corresponding proportion of patients (%) who were frequent ED users in the following year. Points indicate observed proportion of frequent ED users. The diagonal line shows the slope of the line using linear modeling. The error bars are the 95% confidence interval of the binomial distribution of the observed proportion.

Threshold Selection Table. We explored the relationship between ED use in Years 1 and 2, and explored score thresholds that might be used for care management enrollment (Table 4). For example, if a care management team has the capacity to enroll 80 publicly insured patients then it would use a threshold of three. Among the 79 enrolled, the team can expect about 48 patients (61%) to be frequent ED users. If the care management team has a lower enrollment capacity, then they can select a higher score threshold, i.e. enroll fewer patients. Subsequently, for greater enrollment capacities, a lower score threshold can be picked, thereby enrolling more patients. Based on care management capacity, a healthcare organization can select the number of people they wish to enroll.

Table 4. Threshold selection table for care management enrollment recommendations.

Public Insurance

# ED visits in Year 1 (Score Threshold)	Number to Enroll	"Hits" (pts with ≥ 2 ED visits in Year 2)	PPV	Sensitivity	Total # of ED visits in Year 2	Mean # of ED visits per enrollee in Year 2
6+	20	17	85	14	89	4.5
5	30	24	80	20	109	3.6
4	49	36	73	30	185	3.8
3	79	48	61	39	281	3.6
2	150	68	45	56	391	2.6
1	315	97	31	80	527	1.7
0	757	122	16	100	1026	1.4

Private Insurance

# ED visits in Year 1 (Score Threshold)	Number to Enroll	"Hits" (pts with ≥ 2 ED visits in Year 2)	PPV	Sensitivity	Total # of ED visits in Year 2	Mean # of ED visits per enrollee in Year 2
6+	3	3	100	5	32	10.7
5	6	4	67	7	37	6.2
4	15	7	47	12	41	2.7
3	29	9	31	15	71	2.4
2	85	18	21	31	157	1.8
1	267	39	15	66	291	1.1
0	1808	59	3	100	1906	1.1

CHAPTER FOUR
DISCUSSION AND FUTURE DIRECTIONS

Summary. We found several clinical, demographic and healthcare utilization factors that were significantly associated with future frequent ED use among children with asthma. We developed a predictive model that identified children at high risk for future frequent ED use, based on prior ED visit history and insurance status (Public versus Private). This model performed as well as multivariable models and machine learning techniques. Publicly insured patients who visited the ED four or more times in a year had a 50% or higher probability of frequent ED use in the following year; the threshold for privately insured patients was six or more ED visits per year.

Significance. Several important findings emerged from our work. First, upon bivariate analysis, we found significant associations between frequent ED use and many of the potential predictor variables, including race, ethnicity, borough of New York City where patient lives, insurance type, access to Primary Care Physician, history of high healthcare utilization (ED, inpatient, and outpatient visits), medical complexity, BMI, medical comorbidities, technology dependence, allergies, and use of certain medications such as inhaled rescue drugs and mast cell stabilizers. Many of these results are consistent with previous findings in the literature. For example, among children in California, African Americans and Latinos were 82% and 23% respectively more likely than Whites to visit the ED for asthma symptoms.[13] As another example, a study that looked at the use of health services by insurance status among children with asthma from seven New England hospitals found that publicly

insured children used the ED more frequently than privately insured children.[43] Several studies have demonstrated that there are increases in asthma-related ED visits with decreasing SES.[13,43] One explanation is that regions of lower income are more susceptible to high levels of aeroallergens, thereby increasing the likelihood of asthma-related ED visits.[44] Finally, among children admitted to a New England hospital with a primary diagnosis of asthma, a history of previous ED visits was a significant predictor of subsequent ED visits.[12]

Second, despite the large number of significant bivariate associations, we built a predictive model that only required two variables. It is unclear why there was little additional predictive information from the remaining variables. One possible explanation could be that some of these variables are collinear with ED use; indirectly reflected in patient's health utilization patterns, and therefore did not discretely appear in the final models.

Third, although many of the bivariate associations were not useful for predictions, they do provide valuable insight into the kinds of interventions that might be effective for high ED utilization among children with asthma. Some of these factors such as allergies, obesity, and medical complexity can be specifically addressed through clinical care interventions like providing allergy medications, or enrolling patients and caregivers in obesity counseling services. Other factors, such as access to a primary care physician, residing in areas of New York City that have high levels of air pollutants such as midtown and downtown Manhattan, and along busy freeways in

the outer boroughs,[45] and problems associated with low socioeconomic status need more than just clinical interventions. For example, patients who don't have access to a primary care physician may benefit from additional assistance from social workers to find free or subsidized clinics. Care managers can look at some of these predictors that emerged significant in our bivariate analysis to anticipate the healthcare needs of this population, i.e. which patients need only clinical care interventions, and which need more than that.

Fourth, our work demonstrates that EHRs can be valuable source of research data. Many studies use discharge and insurance claims data from statewide or national databases.[22] This approach has several strengths: it has lower data collection costs compared to medical record abstraction or surveys, includes information on uninsured patients not available from third party payers, and provides greater reliability than self-reported medical expenditures. However, there are inconsistencies in how providers, government agencies, and payers report specific data elements. Additionally, there may be hospital-specific errors in data reporting that can lead to data quality problems. An alternative approach is to use data from electronic health records (EHRs). In addition to billing codes, EHRs contain detailed clinical information such as test results, severity of illness, behavioral risk factors, and free text.[23] In our study, many significant variables that were extracted from the EHR are not typically reported in administrative data. Examples of such variables are different types of allergies, IgE test results, and psychiatric and medical comorbidities.

Implications for Clinical Care and Care Management. Our results have implications for clinical practice. For example, healthcare providers can review the insurance status and prior ED utilization history of each patient and make recommendations to care management teams. For example, if a publicly insured patient visited the ED four or more times the previous year, he/she has a 50% or higher likelihood of being a frequent ED user, and could be recommended for enrollment into care management. Additionally, we created a score threshold table for publicly and privately insured patients that may be used for care management enrollment based on available resources. The table gives the care management team a benchmark for comparison. Our model cannot identify which of the enrolled patients are going to be frequent ED users -- thus it is imperative to design tailor-made care plans for each of the individuals enrolled. Results from the bivariate analysis can be good indicators of the kinds of needs that patients may have and could be taken into account when creating individualized care plans. Overall, our study provides recommendations to clinicians and care managers about who could potentially be enrolled in care management programs.

Limitations. There were several important limitations to this study. First, despite the high AUC values of the techniques, the positive predictive values and sensitivities were poor.

Second, our data only captures ED visits to New York Presbyterian-Cornell Medical Center. One cannot rule out the possibility of patients going to other

emergency departments closer to their residences at the time of exacerbations or for asthma management in general.[46-48] Data from other hospital EDs were not available to us at the time of this study. Gathering health utilization data from regional health information exchanges may provide a more comprehensive picture of each patient's ED visits and may help predictive models better detect these events, potentially improving predictive accuracies.

Third, we were unable to operationalize certain concepts that typically appear in free text rather than in structured fields. For example, there are no structured fields to indicate whether a patient is homeless or lives in a shelter. Conceptually, not having a stable residence can have repercussions for disease management, and increase ED utilization for care.[49] Successful operationalization of these concepts by using natural language processing could have made the dataset even more robust and helped improve predictive accuracies.

Fourth, the generalizability of these predictions is unclear. Since our source of data is one medical center in New York City, it is unknown how well our model would work in other locations across the city and other geographic regions in the country.

Future Opportunities. There are several future opportunities that follow from our work. First, the results of the predictive modeling should be implemented and the recommendations for care management should be applied in practice Subsequently, the outcomes could be evaluated for future programmatic modifications.

Second, the value of other data sources such as health information exchanges and personal health records should be explored further for predictive modeling. These may help in creating more robust datasets that may improve predictive accuracies.

Third, other healthcare organizations could use patient data from their own centers and perform predictive modeling to see if the results are consistent with ours. This can address the issue of reproducibility, and can also assess the generalizability of our findings.

Fourth, future research can compare the predictability of frequent ED use among pediatric asthma patients with other common chronic diseases among children such as epilepsy, diabetes, and cystic fibrosis.[50]

Finally, use of predictive modeling in healthcare must overcome the 'impactibility' problem.[51] Even if sophisticated algorithms can accurately identify high-risk patients for readmissions, the extent to which providers can successfully intervene is unclear because of current disease management approaches and resource constraints.[52] Disease management organizations may prioritize managing patients more likely to respond to preventive care, such as patients with ambulatory case sensitive conditions, rather than extremely high-risk patients for whom preventing rehospitalization is unlikely because of factors such as disease severity, mental health diagnoses, severe addictions, demographic characteristics, etc.[51,52] Assessing

efficiencies of these approaches can provide novel insights into how predictive modeling can be better used to create greater impact in healthcare.

Conclusion. We found that publicly insured patients who use the ED four or more times in the current year have a high likelihood of continued frequent ED use (≥ 2 times) in the following year. Privately insured patients who use the ED six or more times in the current year have continued high ED utilization in the following year. Our two variable model can help identify a cohort of patients who may benefit from care management interventions.

BIBLIOGRAPHY

1. Asthma & Children Fact Sheet. American Lung Association. http://www.lung.org/lung-disease/asthma/resources/facts-and-figures/asthma-children-fact-sheet.html. Accessed August 1, 2015.

2. Asthma. Centers for Disease Control and Prevention. http://www.cdc.gov/nchs/fastats/asthma.htm. Accessed August 3, 2015.

3. Trends in Asthma Prevalence, Health Care Use, and Mortality in the United States, 2001-2010. Centers for Disease Control and Prevention. http://www.cdc.gov/nchs/data/databriefs/db94.htm. Accessed August 3, 2015.

4. Asthma in the US. Centers for Disease Control and Prevention. http://www.cdc.gov/vitalsigns/asthma/. Accessed August 1, 2015.

5. Pearson WS, Goates SA, Harrykissoon SD, Miller SA. State-Based Medicaid Costs for Pediatric Asthma Emergency Department Visits. *Prev Chronic Dis.* 2014;11: 140139.

6. Szefler SJ, Zeiger RS, Haselkorn T, et al. Economic burden of impairment in children with severe or difficult-to-treat asthma. *Ann Allergy Asthma Immunol*. 2011;107(2):110-119.e1.

7. Trachtenberg AJ, Dik N, Chateau D, Katz A. Inequities in ambulatory care and the relationship between socioeconomic status and respiratory hospitalizations: a population-based study of a Canadian city. *Ann Fam Med*. 2014;12(5):402-7.

8. Enard KR, Ganelin DM. Reducing preventable emergency department utilization and costs by using community health workers as patient navigators. *J Healthc Manag*. 2013;58(6):412-27.

9. Sweeney L, Halpert A, Waranoff J. Patient-centered management of complex patients can reduce costs without shortening life. *Am J Manag Care*. 2007;13(2):84-92.

10. McCarthy D, Cohen A, Johnson MB. Gaining ground: care management programs to reduce hospital admissions and readmissions among chronically ill and vulnerable patients. *Case Study Series Synthesis The Commonwealth Fund.* 2013.

11. Billings J, Blunt I, Steventon A, Georghiou T, Lewis G, Bardsley M. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open*. 2012;2(4).

12. Tolomeo C, Savrin C, Heinzer M, Bazzy-asaad A. Predictors of asthma-related pediatric emergency department visits and hospitalizations. *J Asthma*. 2009;46(8):829-34.

13. Wright K. Disparities and predictors of emergency department use among California's African American, Latino, and White children, aged 1-11 years, with asthma. *Ethn Dis*. 2009;19(1):71-7.

14. Steppuhn H, Langen U, Scheidt-nave C, Keil T. Major comorbid conditions in asthma and association with asthma-related hospitalizations and emergency department admissions in adults: results from the German National Health Telephone Interview Survey (GEDA) 2010. *BMC Pulm Med*. 2013;13:46.

15. Ferris TG, Crain EF, Oken E, et al. Insurance and quality of care for children with acute asthma. *Ambul Pediatr*. 2001;1(5):267-74.

16. Schatz M, Zeiger RS, Vollmer WM, et al. The controller-to-total asthma medication ratio is associated with patient-centered as well as utilization outcomes. *Chest*. 2006;130(1):43-50.

17. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA*. 2011;306(15):1688-98.

18. Shulan M, Gao K, Moore CD. Predicting 30-day all-cause hospital readmissions. *Health Care Manag Sci*. 2013;16(2):167-75.

19. Yu S, van Esbroeck A, Farooq F, Fung G, Anand V, Krishnapuram B. Predicting Readmission Risk With Institution Specific Prediction Models. In: *Proceedings of the IEEE International Conference on Healthcare Informatics 2013.* 2013. p. 551-556.

20. Mather JF, Fortunato GJ, Ash JL, Davis MJ, Kumar A. Prediction of pneumonia 30-day readmissions: a single-center attempt to increase model performance. *Respir Care*. 2014;59(2):199-208.

21. Natale J, Wang S, Taylor J. A Decision Tree Model for Predicting Heart Failure Patient Readmissions. In: Krishnamurthy A, Chan WKV. (eds.) *Proceedings of the 2013 Industrial and Systems Engineering Research Conference.* 2013.

22. Ohno-machado L. Realizing the full potential of electronic health records: the role of natural language processing. *J Am Med Inform Assoc*. 2011;18(5):539.

23. Why Predictive Modeling in Healthcare Requires A Data Warehouse. Health Catalyst. https://www.healthcatalyst.com/wp-content/uploads/2014/02/Why-predictive-modeling-healthcare-requires-a-data-warehouse.pdf. Accessed January 28, 2015.

24. Moorman JE, Rudd RA, Johnson CA, et al. National surveillance for asthma--United States, 1980-2004. *MMWR Surveill Summ*. 2007;56(8):1-54.

25. Simon TD, Cawthon ML, Stanford S, et al. Pediatric medical complexity algorithm: a new method to stratify children by medical complexity. *Pediatrics*. 2014;133(6):e1647-54.

26. Hunt KA, Weber EJ, Showstack JA, Colby DC, Callaham ML. Characteristics of frequent users of emergency departments. *Ann Emerg Med*. 2006;48(1):1-8.

27. Byrne M, Murphy AW, Plunkett PK, Mcgee HM, Murray A, Bury G. Frequent attenders to an emergency department: a study of primary health care use, medical profile, and psychosocial characteristics. *Ann Emerg Med*. 2003;41(3):309-18.

28. Ledoux Y, Minner P. Occasional and frequent repeaters in a psychiatric emergency room. *Soc Psychiatry Psychiatr Epidemiol*. 2006;41(2):115-21.

29. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction, 2nd ed. New York, NY: Springer, 2009.

30. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.

31. Rpart: Recursive Partitioning. R Package version 4.1-4. [computer program] 2015.

32. Liaw A, Wiener M. Classification and Regression by randomForest. *R News*. 2002;2:18-22.

33. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R Package version 1.6-2 [computer program] 2014.

34. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

35. Warnes GR, Bolker B, Gorjanc G, et al. gdata: Various R programming tools for data manipulation. R package version 2.13.3. 2014.

36. Milborrow S. rpart.plot: Plot rpart Models. R package version 1.5.2. 2015.

37. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940-1.

38. Kuhn M. caret: Classification and regression training. R package version 6.0-47.

39. Wickham H. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

40. Dowle M, Short T, Lianoglou S. data.table: Extension of data.frame. R package version 1.9.4. 2014.

41. Lumley T. leaps: regression subset selection. R package version 2.9. 2009.

42. Xavier R, Turck N, Hainard A, et al. pROC: an open-source packager for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.

43. Ortega AN, Belanger KD, Paltiel AD, Horwitz SM, Bracken MB, Leaderer BP. Use of health services by insurance status among children with asthma. *Med Care*. 2001;39(10):1065-74.

44. Heffernan PM. Socioeconomic status, asthma, and emergency department use in Ontario, Canada. [master's thesis]. St John's, Canada: Memorial University of Newfoundland; 2012.

45. The New York City community air survey. New York City Department of Health and Mental Hygiene. http://www.nyc.gov/html/doh/downloads/pdf/eode/comm-air-survey-report.pdf. Accessed August 1, 2015.

46. Shapiro JS, Johnson SA, Angiollilo J, Fleischman W, Onyile A, Kuperman G. Health information exchange improves identification of frequent emergency department users. *Health Aff (Millwood)*. 2013;32(12):2193-8.

47. Bourgeois FC, Olson KL, Mandl KD. Patients treated at multiple acute health care facilities: quantifying information fragmentation. *Arch Intern Med*. 2010;170(22):1989-95.

48. Finnell JT, Overhage JM, Grannis S. All health care is not local: an evaluation of the distribution of Emergency Department care delivered in Indiana. *AMIA Annu Symp Proc*. 2011;2011:409-16.

49. Ku BS, Scott KC, Kertesz SG, Pitts SR. Factors associated with use of urban emergency departments by the U.S. homeless population. *Public Health Rep*. 2010;125(3):398-405.

50. Torpy JM, Campbell A, Glass RM. JAMA patient page. Chronic diseases of children. *JAMA*. 2010;303(7):682.

51. Lewis GH. "Impactibility models": identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. *Milbank Q*. 2010;88(2):240-55.

52. Wharam JF, Weiner JP. The promise and peril of healthcare forecasting. *Am J Manag Care.* 2012;18(3):e82-5.