Using Daily Progress Note Data to Predict Discharge Date from the Neonatal Intensive Care Unit

By

Michael William Temple

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Informatics

August, 2015

Nashville, Tennessee

Approved:

Christoph U. Lehmann, M.D.

Kevin B. Johnson, M.D., M.S.

Daniel Fabbri, Ph.D.

William Gregg, M.D., M.S., M.P.H.

DEDICATION

To my amazingly supportive wife, Shelley

and

To my two marvelous children, Brendan and Gabby.

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support of Vanderbilt University and the National Library of Medicine (training grant 5T15LM007450).

I am grateful for all of the people I have had the pleasure to work with over the past several years. All the members of my thesis committee have taught me valuable lessons about scientific research and the importance of making the work meaningful. I would especially like to thank the chair of my committee Dr. Christoph Lehmann for his guidance in research direction and insights into producing quality manuscripts. Dr. Kevin Johnson has been a friend and mentor and I appreciate his willingness to take a chance on a more "non-traditional" student.

Finally, none of this would have been possible without the unwavering support of my family. My wife, Shelley, and children, Brendan and Gabby, have been unbelievably supportive and understanding as I pursued this goal. I am forever in their debt.

TABLE OF CONTENTS

LIST OF TABLES

Chapter II

Chapter III

LIST OF FIGURES

Chapter II

Chapter III

CHAPTER I

INTRODUCTION

Research Motivation

The environment for delivering healthcare is becoming more challenging. Hospitals are faced with economic constraints and decreasing capacity as they try to continue to improve the quality of care delivered. To increase the efficiency of care delivered, hospitals have begun to focus resources on the management of patient flow within the hospital and patient length of stay (LOS).

Improving efficiency of care and decreasing the LOS have a real impact on the financial performance of the hospital. Hospital reimbursement is often provided in a framework based on a Diagnostic Related Group (DRG). In this framework, hospitals are given a lump sum payment to manage the needs of a patient with a particular diagnosis. If the payment is meant to cover an illness that usually requires three days of hospitalization and the patient can be discharged in two, then the hospital benefits by reducing cost through reduced services provided (such as nursing care, supplies, medications, food) and is able to make the bed available to the next patient. On the other hand, if the patient remains in the hospital for five days, the hospital is not paid any additional monies, has to absorb the added costs, and is unable to fill the bed with another patient.

One of the areas with the highest daily cost for the hospital is the intensive care unit. For a pediatric hospital this would include the pediatric intensive care unit (PICU) and the neonatal intensive care unit (NICU). These two areas are also at the center of patient flow for pediatric hospitals – intersecting with the Emergency Department, Operating Rooms as well as the regular wards. Managing the flow, length of stay, and efficient use of resources as patients are moved among these interdependent, complex systems can have a significant financial impact for the hospital organization.

The average length of stay (LOS) in the NICU at Monroe-Carell Children's Hospital at Vanderbilt University Medical Center (VUMC) has been increasing over the past four years. In 2010 the average LOS was 21 days. In 2013, that figure was 26 days. The increased LOS has negative financial implications for the institution since most payments are fixed DRG payments based on the underlying clinical problems. Additionally, increased length of stay can lead to additional complications, such as life-threatening infections, for the infants in the unit.

The NICU population has a wide array of diseases with varying complexity and LOS. Disorders can range from an infant with a severe cardiac anomaly requiring several cardiac surgeries to a premature infant with mild respiratory issues to a term infant with presumed infection. Adding to the complexity is the need for social work involvement and a vast amount of parent education and training regarding numerous topics including feeding schedules, medication usage, and home medical equipment instruction. Some patients may be in the NICU for a number of months and their needs can shift from critical care to primary care requiring the need for vaccinations and developmental screenings. Additionally, the NICU at VUMC is spread over four different locations separated by a quarter of a mile in the hospital with four different medical teams that change their attending physician every two weeks.

The discharge dates tend to be a moving target in part because of differences in discharge criteria among attending physicians, who change service responsibility every other Monday. Other potential delays in discharge stem from lack of training for the infant's parents, incomplete screening tests, lack of required home equipment, complications involving child protective services, lack of parental means of transportation, or deterioration of the patient's status. Frequently social issues like exposure to substances in utero and the requirement to be cleared or placed into foster care cause delays in discharge. A lot of the staff members that perform parent education and training are not available in the evening or on the weekends. With parents who are employed, however, the evening and weekends are the most likely times that they will be in the hospital and available to receive their training. These extraneous factors are not related to the patient's medical condition and the infant's discharge can be delayed several days because of these factors.

All of the above factors – variability in patient complexity, availability of staff and parents for training, attending physician preferences, multiple locations, and lack of comprehensive informatics tools – may result in delay in discharge, which makes predicting the discharge of NICU patients very difficult. Subsequently, the forecasting of the census for the unit and the necessary staffing becomes quite challenging.

Since infants are most frequently discharged home directly from the NICU (and not transferred to another floor of the hospital prior to discharge) a key issue for this project is the idea of "medically ready for discharge". Many times in the NICU, the patient is ready to be discharged home from a medical standpoint, but other social or discharge planning roadblocks remain that prevent the patient from going home. Custody issues, parent education and arranging home-going medical equipment are the most common causes of these extended lengths of stay. By predicting which patients will be medically ready for discharge in the upcoming week, the hope is that the social or discharge planning issues can be resolved prior to the infant being ready for discharge. This will decrease the length of stay for these infants.

Specific Aim # 1: Create a model to predict when NICU patients will be medically ready for discharge.

The focus of this project is not to predict LOS from time of admission. This project will use clinical data extracted from the daily progress notes and attempt to predict which patients will be medically ready for discharge in the next 10 days. The prediction model will be created using a Random Forest in combination with the extracted clinical data. Identification of patients who will be medically ready for discharge will provide enough lead-time to the clinical staff to resolve any non-medical issues that could potentially delay the discharge for a patient. This will allow the patient to be discharged as soon as they are medically ready.

Specific Aim # 2: Identify the most important clinical features that have the greatest impact on the accuracy of the discharge prediction model.

Once the prediction model has been created, analysis of the performance of clinical features in the model will be examined to determine which ones are the most critical for predictive accuracy. It is highly likely that a few critical clinical features will be responsible for a large part of the predictive accuracy of the model. Some features may be more difficult to extract than others and the consistency in documentation may make some features less reliable. Identifying the most critical features could allow for simpler and more consistently accurate models.

Specific Aim # 3: Once a predictive model has been created, identify which patients performed poorly in the model and the reason for the poor performance.

In order to refine and improve on the prediction model, identification of poorly performing patients and the reasons for that poor performance will be crucial. It is likely that the first iterations of the model will miss some important features for some patients. Identifying poor performing patients and devising a method to discover the reasons for that poor performance will allow for further refinement and improvement of the predictive model.

The first manuscript in this thesis will focus on the first two aims, and the third aim will be addressed in the second manuscript.

CHAPTER II

USING DAILY PROGRESS NOTE DATA TO PREDICT DISCHARGE DATE FROM THE NEONATAL INTENSIVE CARE UNIT *

Michael W. Temple[1], MD, Christoph U. Lehmann[1, 2], MD, Daniel Fabbri[1], PhD

Affiliations: [1]Department of Biomedical Informatics, [2]Department of Pediatrics Vanderbilt University, Nashville, TN.

Address correspondence to: Michael Temple, Department of Biomedical Informatics, Vanderbilt University School of Medicine, 2525 West End, Suite 1475, Nashville, TN 37203-8390, [michael.w.temple@vanderbilt.edu], 615-936-1068.

Short title: Predicting Discharge Date from the NICU.

Abbreviations: AUC – Area under the Curve, CART -- Classification And Regression Trees, DTD – Days to Discharge, GI – Gastrointestinal, LOS – Length of Stay, NICU – Neonatal Intensive Care Unit, NS – Neurosurgery, RF – Random Forest.

Key Words: Intensive Care Units, Neonatal; Area Under Curve; Patient Discharge; ROC Curve

Funding Source: National Library of Medicine Training Grant 5T15LM007450-13.

Financial Disclosure: Dr. Lehmann serves in a part-time role at the American Academy of Pediatrics. He also received royalties for the textbook Pediatric Informatics, and travel funds from the American Medical Informatics Association, the International Medical Informatics Association and the World Congress on Information Technology.
Dr. Fabbri has an equity interest in Maize Analytics, LLC. Dr. Temple has no financial disclosures.

Conflict of Interest: The authors have no conflicts of interest to disclose.

What's Known on This Subject: Discharging patients from the NICU require coordination and may be delayed for non-medical reasons. Predicting when patients will be "medically ready" for discharge can avoid these delays and result in cost savings for the hospital.

What This Study Adds: We developed a supervised machine learning approach leveraging real-time patient data from the daily neonatology progress note to predict when patients will be medically ready for discharge.

*** Manuscript accepted for publication by *Pediatrics*. Publication Pending.**

Abstract

Background and Objectives
Discharging patients from the Neonatal Intensive Care Unit (NICU) may be delayed for non-medical reasons including the need for medical equipment, parental education, and children's services. We describe a method to predict and identify patients that will be medically ready for discharge in the next 2-10 days – providing lead-time to address non-medical reasons for delayed discharge.

Methods
A retrospective study examined 26 features (17 extracted, 9 engineered) from daily progress notes of 4,693 patients (103,206 patient-days) from the NICU of a large, academic children's hospital. A matrix was constructed using these features and the days to discharge (DTD). Patients were classified as premature, cardiac, GI surgery, and/or neurosurgery based on ICD-9 codes. A supervised machine learning approach using a Random Forest defined the most important features and created a discharge prediction model.

Results
Three of the four sub-populations (Premature, Cardiac, GI surgery) and all patients combined performed similarly at 2, 4, 7, and 10 DTD with AUC ranging from 0.854-0.865 at 2 DTD and 0.723-0.729 at 10 DTD. Neurosurgery patients performed worse at every DTD measure scoring 0.749 at 2 DTD and 0.614 at 10 DTD. This model was also able to identify important features and provide "rule-of-thumb" criteria for patients close to discharge. Using DTD equal to 4 and 2 features (oral percentage of feedings and weight) we constructed a model with an AUC of 0.843.

Conclusion
Using clinical features from daily progress notes provides an accurate method to predict when NICU patients are nearing discharge.

Introduction

Approximately four million babies are born every year in the United States and about 11% [~440,000] of those are born prematurely.[1] Caring for infants in the Neonatal Intensive Care Unit (NICU) poses a significant financial burden to the health care system with an estimated total cost of 26 billion dollars.[1] The cost per day of NICU care can be several thousand dollars; therefore discharging these infants as soon as they are medically ready is critical to controlling expenditures.

Delayed discharge of hospitalized patients who are medically ready is a common occurrence often linked to dependency and the need to provide post-discharge services.[2] In elderly patients, difficulties in coordinating post-discharge services, lack of anticipation of discharge, and absence of caregivers at home were associated with delayed discharge of medically ready patients.[3] Similarly, discharging a patient from the NICU usually requires a great deal of coordination. Neonates discharged from the NICU are prime examples of patients with dependencies (on parents and caregivers) and significant post-discharge needs like primary care, specialists, physical and speech therapy, neonatal follow-up appointments, home equipment services, and home nursing. In cases of intra-uterine drug exposure, discharge is often dependent upon Child Protective Services approval. Parents have to demonstrate their ability to operate medical equipment, to administer home medication, and to feed and care for their medically fragile infant. In addition, a number of services must be scheduled around the time of discharge such as hearing screens, car seat tests, immunizations, repeat state screens, and eye exams. All of these requirements can delay the discharge of a patient who is medically ready and, consequently, unnecessarily increase the cost of hospitalization.

The goal of this project is to build a predictive model to identify those patients who are close to discharge from a medical perspective so staff can be alerted to impending discharges. This will allow the non-medical factors to be addressed in advance to ensure the patient's discharge will not be delayed.

Almost all previous studies attempt to predict length of stay (LOS) using clinical and diagnostic information at (or near) the time of admission.[4-7] While it is important to pursue LOS prediction to understand total hospitalization costs, these methods lack sufficient clinical context to accurately predict the discharge date. Instead, the focus of this research project is to identify, based on the most recent clinical data, which NICU patients will likely be discharged home in the next 2-10 days. Our methodology predicts the upcoming discharge date – not the LOS from time of admission.

In order to prevent delayed discharge, three questions will be answered. First, can the discharge date for a NICU patient be accurately predicted? Second, what combinations of clinical data improve predictive accuracy? Lastly, are there simple, "rule-of-thumb" factors that are responsible for a substantial fraction of the prediction accuracy?

Related Work

Because of the potential impact on cost savings, predicting the LOS for NICU patients has been well studied. Most of the following prediction methods were performed at or near the time of admission. Powell et al. found gestational age, low birth weight, and respiratory difficulties to be most predictive of LOS.[8] Bannwart et al. developed two models to predict the LOS for patients in the NICU.[9] The first model only considered risk factors present in the first three days of life, while the second model used factors present during the entire hospitalization.

Despite the use of models incorporating multiple diagnostic factors at the time of admission and during the hospitalization, the accuracy of these models varied significantly making LOS prediction difficult. Lee et al. studying the Canadian NICU Network found that "significant variation in NICU practices and outcomes was observed despite Canada's universal health insurance system".[10] Lee et al. using data from "The California Perinatal Quality Care Collaborative" reported "wide variance in LOS by birth weight, gestational age, and other factors".[11]

In 2012, Levin et al. described a real-time model to forecast LOS in a PICU using physician orders from a Provider Order Entry system.[12] This model used physician orders (not diagnostic data) to provide a cumulative probability of discharge from the PICU over the next 72 hours. Counts of medications by administration route (injected, infused, or enteral) were more significant in predicting discharge from the PICU than the types of medication the patient received. Activity, diet (regular diet vs. parenteral nutrition) and mechanical ventilation orders were highly predictive of remaining in the PICU over the next 72 hours.

It was our hypothesis that using a real-time data source that reflects orders, physiologic data, and diagnostic information will allow for improved NICU discharge prediction.

In contrast to LOS models that are performed at the time of admission, our model is updated daily with the most recent progress note data. The calculated probability of discharge may, in the future, be displayed in the electronic medical record.

Methods

Patients and Setting

We conducted a retrospective study of all patients admitted to the NICU at a large academic medical center from June 2007 to May 2013.

Exclusion Criteria

All patients admitted to the NICU were considered for the study. Patients who were back-transferred to another facility or who died during the course of their NICU hospitalization were excluded from the analysis. Also excluded from the analysis were patients with any missing daily neonatology progress notes.

Data Collection and Extraction

A large database containing all of the daily progress notes written by neonatology attending physicians was made available to the investigators. The data from the progress notes were in a semi-structured text format that was extracted using regular expressions in Python (version 2.7.3) and SQL. In addition, these data were cross-referenced with the enterprise data warehouse in order to obtain basic patient information such as date of birth and ICD-9 codes used for billing during the hospitalization.

Feature Descriptions

The clinical features used in our model fell into four main categories: quantitative, qualitative, engineered, and derived sub-populations. Thirteen features were obtained directly from data contained within the daily progress notes. These extracted features were classified as quantitative (values fell within a range) and qualitative (assigned a value of 0 or 1). Nine features were engineered from the extracted data. These engineered features do not actually exist as data in the progress note but were derived from the extracted data. For example, progress notes contain information on the number of apnea and bradycardia events (A&B's) in the last 24 hours. The engineered feature from these data was the number of days since the last A&B.

Additionally, a neonatologist (CU Lehmann) reviewed 138 of the most frequently occurring ICD-9 codes in the NICU patient population to categorize patients into 4 sub-populations: Prematurity, Cardiac disease, Gastrointestinal (GI) Surgical disease, and Neurosurgical (NS) disease (please see Appendix 1 for a list of ICD-9 codes and categories). A single patient could belong to one, many, or none of the sub-populations. Table 1 contains a list of all features used in the model.

Table 1. Features used in the Predictive Model

Quantitative Features (Units)	Qualitative Features (Units)	Engineered Features (Units)	Sub-Population Features
Weight (kg)	On Infused Medication (Y/N)	Number of Days Since Last A&B Event(days)	Premature (Y/N)
Birth Weight (kg)	On Caffeine (Y/N)	Number of Days Off Infused Medication (days)	Cardiac Surgery (Y/N)
Apnea and Bradycardia (A&B) Events (number)	On Ventilator (Y/N)	Number of Days Percent of Oral Feeds > 90% (days)	GI Surgery (Y/N)
Amount of Oral Feeds (ml)		Number of Days Off Ventilator (days)	Neurosurgery (Y/N)
Amount of Tube Feeds (ml)		Number of Days Off Oxygen (days)	
Percentage of Oral Feeds (%)		Number of Days Off Caffeine (days)	
Gestational Age (weeks)		Total Feeds (Oral + Tube Feeds) (ml)	
Gestational Age at Birth (weeks)		Ratio of Weight to Birth Weight	
Day of Life (days)		Amount of Oral Feeds / Weight (ml/kg/day)	
Oxygen (per liter)			

Matrix Generation

All of the extracted data, sub-population categories, engineered features, and days to discharge (DTD) were inserted into a matrix. Each row represented data for one hospital day for a specific patient. If a row contained missing data in any field, the entire row was excluded from the final matrix.

Since the matrix is constructed using historical data, the outcome of interest (discharge date) is known. The DTD column contains the number of hospital days until the patient is discharged. For example, if the patient was discharged on March 15, the row of the matrix containing patient features for March 10 would have a DTD of 5 (Figure 1).

Data Analysis

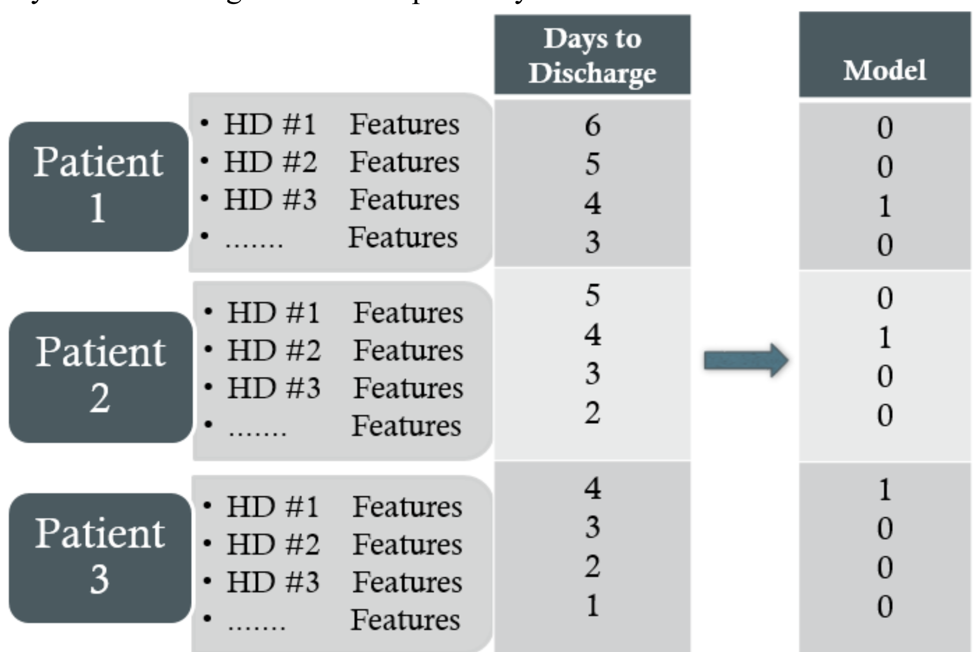
A supervised machine learning approach using a Random Forest (RF) classifier in Python's Sci-kit Learn module (version 0.15.2)[13] was used to analyze the data, engineer important features, and build a predictive model. A RF constructs many binary decision trees that branch based on randomly chosen features. The RF in Sci-kit Learn uses an optimized Classification And Regression Trees (CART) algorithm for constructing binary trees using the input features and values that yield the largest information gain at each node. The Sci-kit Learn package allows for the selection of either the gini impurity or entropy algorithms to determine feature importance. These algorithms performed similarly and we chose to use gini impurity because it is slightly more robust to misclassifications. We ran the models using many different combinations of parameters and the best performing models used a RF with 100 trees, maximum tree depth of 10 and a minimum of 200 samples per split.

Models were trained using different combinations of sub-populations (all patients, premature, cardiac, GI surgery, and neurosurgery), DTD (2, 4, 7, and 10 days) and number of features (any combination of features from 2 to all 26).

Training Vector

In order to train our model, we converted the number of "Days to Discharge" variable into a binary outcome variable based on the number of days we were trying to model. For example, if we were training the model to predict when patients were four days from discharge, all values in the model where the DTD was not equal to four were set to "0". The rows in which the number of DTD was four, were set to "1" (Figure 1). This same process was followed for 2, 7, and 10 DTD.

Figure 1. Example data matrix construction. This provides an example if trying to model four days until discharge. HD = Hospital Day



Cross Validation

Each time a model was run, half of the patients (and all their associated daily rows) were randomized into a training set and the other half were assigned to the testing set. Since each patient provides only a single DTD, halving the data provided both testing and training sets an adequate number of the DTD of interest. To achieve small enough standard deviations, the patients were randomized a total of five times for each model and the area under the curve (AUC) for the receiver operating characteristic (ROC) curve was obtained for the testing set. The reported AUC is the average of the five AUC's obtained after each round of randomization.

Additionally, each time a model was run, the features used in the model were ranked in order of importance.

Model Generation

We ran the model for all patients and for each sub-population to determine how well the model performed, to decide the most important features for each group, and to determine if different features had a greater impact on certain patient populations. Finally the most important features at 2, 4, 7, and 10 days to discharge were evaluated to determine if the most important features changed as a patient was getting closer to discharge.

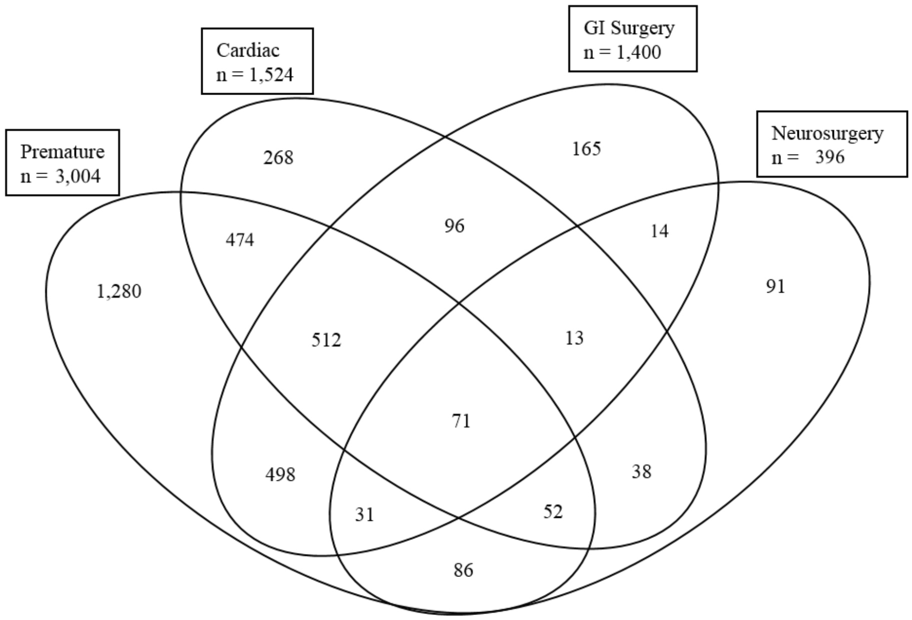
IRB Approval

The Institutional Review Board of Vanderbilt University approved this study.

Results

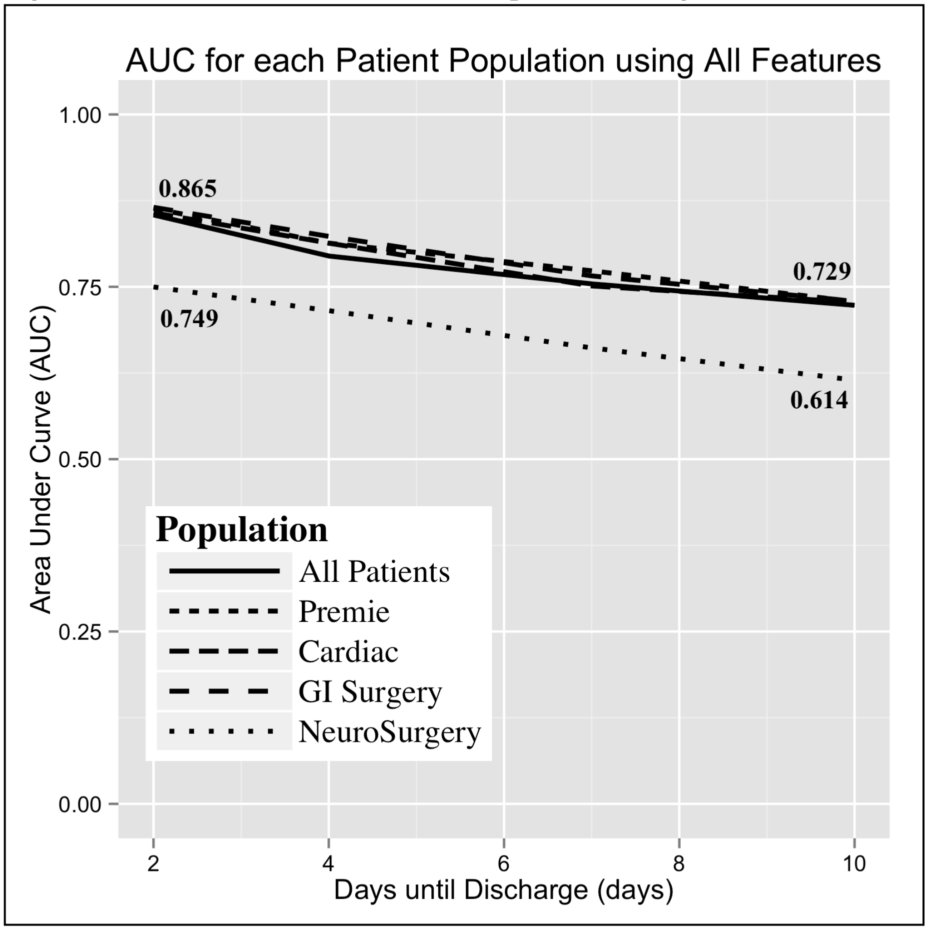
The initial database consisted of 6,302 patients (116,299 hospital days) admitted to the NICU between June 2007 and May 2013. There were 256 (4%) deaths during this time period. A total of 1,154 (18%) patients were excluded because the database did not contain physician progress notes for every day of the hospital course. There were 199 (3%) patients back-transferred to other NICU's in the region. The final matrix consisted of 4,693 (74%) unique patients accounting for 103,206 (89%) hospital days with a mean LOS of 30 days. A total of 3,689 (79%) patients were categorized into one or more sub-populations based on ICD-9 codes; the other 1,004 (21%) patients did not have an ICD-9 code that matched our criteria (Figure 2).

Figure 2. Distribution of patients in each sub-population



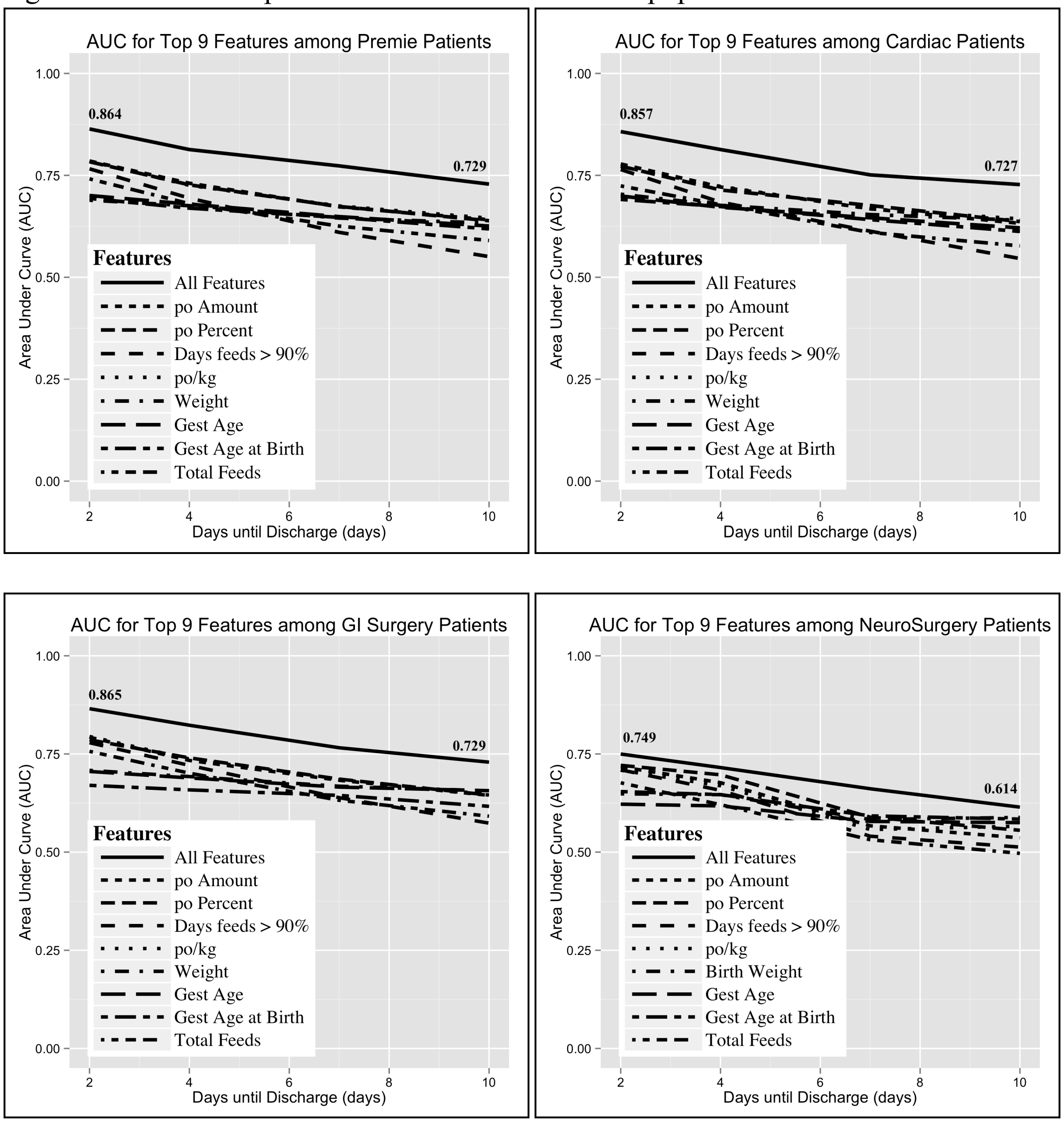
The average AUC for the model using all 26 features for all patients and each patient sub-population is shown in Figure 3. Three of the four sub-populations (Premature, Cardiac, GI surgery) and all patients combined performed very similarly at 2, 4, 7, and 10 DTD with AUC scores ranging from 0.854-0.865 at 2 DTD and 0.723-0.729 at 10 DTD. The Neurosurgery sub-population performed worse at every DTD measure scoring 0.749 at 2 DTD and 0.614 at 10 DTD (Figure 3). Using five-fold cross-validation provided a sufficiently narrow standard deviation range for AUC's of approximately 0.005-0.01.

Figure 3. AUC for each Patient Sub-Population using All Features



The nine most predictive features for each sub-population were very similar and their plots are shown in Figure 4. In each sub-population, the combination of all features performed better than any single feature alone. Once again the poorest performing sub-population included the neurosurgery patients.

Figure 4. The 9 most predictive features for each sub-population



* A single patient may be represented in more than 1 sub-population.

In addition to analyzing the most important features for each sub-population, we also explored the best performing features by the DTD. For each DTD (2, 4, 7, 10 days) the top 20 features in order of importance are shown in Table 2. The combination of all features performed best at each DTD, and model performance improved as patient moved closer to discharge.

Table 2. The top 20 features in order of importance for all patients for all days until discharge

2 Days Until Discharge		4 Days Until Discharge		7 Days Until Discharge		10 Days Until Discharge	
Feat	AUC	Feat	AUC	Feat	AUC	Feat	AUC
All	0.854	All	0.795	All	0.754	All	0.723
% of Oral Feeds	0.766	% of Oral Feeds	0.704	Amt Oral Feeds	0.649	% of Oral Feeds	0.623
Amt Oral Feeds	0.764	Amt Oral Feeds	0.703	Gest Age Birth	0.647	Amt Oral Feeds	0.620
# Days Oral % > 90%	0.753	Amt Oral Feeds / Weight	0.700	Amt Oral Feeds / Weight	0.646	Amt Oral Feeds / Weight	0.620
Amt Oral Feeds / Weight	0.750	# Days Oral % > 90%	0.681	% of Oral Feeds	0.646	Gest Age	0.617
Total Feeds	0.720	Gest Age Birth	0.678	Birth Weight	0.632	Weight	0.617
Gest Age Birth	0.707	Gest Age	0.673	Weight	0.632	Gest Age Birth	0.609
Birth Weight	0.698	Birth Weight	0.672	Gest Age	0.631	Birth Weight	0.607
Gest Age	0.698	Weight	0.667	# Days off Caffeine	0.610	On Caffeine	0.594
Weight	0.690	Total Feeds	0.652	On Caffeine	0.605	# Days off Caffeine	0.592
# Days off Caffeine	0.643	# Days off Caffeine	0.630	GI Surgery	0.594	Total Feeds	0.569
GI Surgery	0.637	GI Surgery	0.622	Total Feeds	0.590	GI Surgery	0.566
# Days off Vent	0.624	On Caffeine	0.608	# Days Oral % > 90%	0.589	# Days off Oxygen	0.560
# Days off Infused Meds	0.620	# Days off Vent	0.605	Cardiac	0.563	On Oxygen	0.548
Ratio Weight / Birth Weight	0.613	# Days off Infused Meds	0.594	# Days off Vent	0.562	On Vent	0.543
On Caffeine	0.613	Ratio Weight / Birth Weight	0.592	Ratio Weight / Birth Weight	0.561	Cardiac	0.542
# Days off Oxygen	0.609	Day of Life	0.587	# Days off Oxygen	0.558	# Days off Vent	0.537
Day of Life	0.604	Cardiac	0.582	# Days off Infused Meds	0.555	# of A&B	0.535
# Days No A&B	0.601	# Days off Oxygen	0.581	Day of Life	0.547	# Days Oral % > 90%	0.534

Discussion

We were able to use data from daily progress notes to predict impending discharge accurately from the NICU. Our model improved as more clinical information was included and its prediction improved as the DTD became smaller (closer to discharge date). Three of the four sub-populations as well as all patients combined performed very similarly. The one population on which the model consistently underperformed was the neurosurgery population. First, the neurosurgery population was the smallest cohort by far and therefore the model may not have had enough patients on which to adequately train. Second, it could also suggest that the neurosurgery population may be very different clinically than the other patients seen in the NICU and their readiness for discharge may not be captured in the features extracted for this model.

When breaking the most important features down by each sub-population and DTD, the features remained surprisingly consistent across the populations and DTD. This was unexpected as we felt that different sub-populations of patients with different medical conditions would have different features that were important for discharge prediction. The top features centered on various feeding metrics, gestational age, and weight. Surprisingly, none of the metrics involving infused medications, caffeine use, A&B's, or oxygen usage had a significant impact on the predictive power of the model.

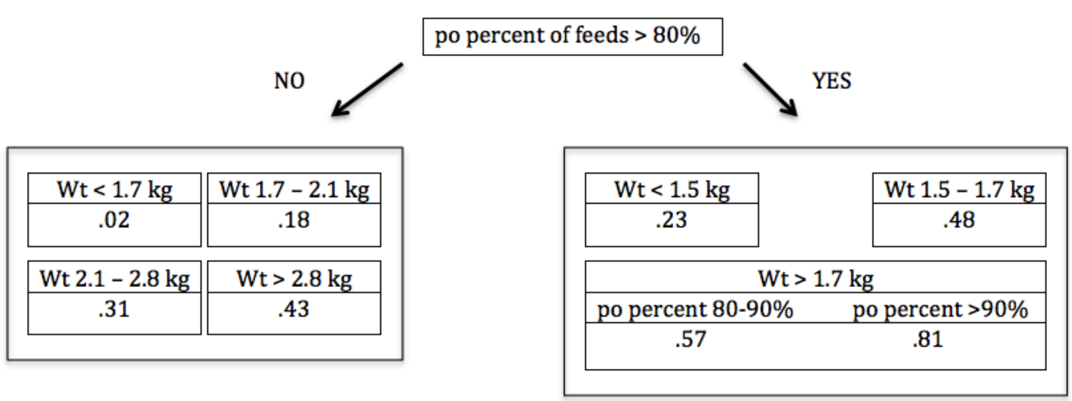
Two interesting features are worth discussing. First, the percentage of oral feeds (e.g., oral amount divided by the oral amount plus the tube fed amount) was the top, or near the top, performing feature across populations and DTD. As an example, using this feature alone gives an AUC score of 0.766 at 2 DTD. The second best feature was the engineered feature of the number of days with oral feedings of greater than 90%. At 10 DTD this feature ranks 20th in importance, but at 2 DTD this feature has advanced to 3rd place. This indicates that consuming the vast majority of their feedings orally instead of by tube is an important predictor of impending discharge.

We used 26 features to predict with a high degree of accuracy which patients will be discharged home in the next 2-10 days. However, it may not always be practical or possible to include all of these features into a decision support tool in order to construct this predictive model to alert staff of impending discharges. One of the beneficial aspects of our approach is the ability to identify and use the most important features to build a scaled down but still highly predictive model.

A few, simple "rule of thumb" models can be created to identify patients who are nearing discharge. As an example, using only two features, a very simple decision tree can be constructed (Figure 5). This tree was created using all patients, two features (oral percentage of feeds and weight), a DTD of four days and a maximum tree depth of three. The first branch of the tree splits the patients into 2 groups based on whether or not their oral percentage of feeds is greater than 80%. Following this path to the right, the next differentiator is based on weight. If the patient weighs less than 1.5 kg, the probability for them to be discharged in the next four days is 0.23 (on a scale of 0-1). If they weigh between 1.5 and 1.7 kg, then their probability for discharge in the next four days is 0.48. If the patient weighs more than 1.7 kg and they take more than 90% of their feeds orally, then they have a 0.81 probability of being discharged in the next four days. The probabilities for discharge in four days for patients at different weights and taking less than 80% of their feeds orally are listed in the left-side branch.

This simple decision tree has an AUC of 0.843. While it is not as accurate as using all features to obtain an AUC of 0.865, it is still an excellent predictor and can be easily calculated at the bedside.

Figure 5. A simple decision tree demonstrating how two features can be used to create a relatively accurate discharge prediction model. The fraction in each cell denotes the probability of discharge in the next four days. This tree has an AUC = 0.843.



It is interesting that all 26 features gives an AUC of 0.865 while using only 2 features can give an AUC 0.843. This result illustrates just how important feeding and weight gain are to the improving health of a neonate.

One possible way to improve our current model performance would be to add more features. The use of trending data (e.g., the average amount of feeding increase over a five day period) could prove to be beneficial. Another consideration for model improvement would be to predict a range of days until discharge (for example, 3-5 days instead of just 4).

Limitations and Next Steps

There are several limitations to this study. First, some of the features used in the model are more difficult to obtain than others, and the ability to extract certain features from commercial electronic medical record systems can be challenging.[14] Second, the data extracted included pediatric and neonatology specific data, which was collected using specific pediatric functionality built into Vanderbilt's electronic health record. These functionalities may not be supported by all electronic health record systems.[15,16] Third, categorizing hospitalized patients based on ICD-9 codes would be difficult since these codes are not usually available until after discharge. However, as the analysis showed, diagnosis categories added surprisingly little to the prediction model. Should, in the future, our model need to differentiate patients, admitting diagnoses could be used. Fourth, our sample could be potentially biased since we did exclude patients if they were missing any progress notes. While a Random Forest does provide techniques to address missing data, we felt that excluding these patients was a conservative and appropriate approach.

We trained the model using actual discharge dates. This limitation worked against us since some of the patients in the data set may have been medically ready for discharge sooner. The model may have performed better if we had been able to determine and adjust for the patients that had delayed discharges for non-medical reasons. Additionally, our model might – once fully implemented – predict discharge too early, which could result in premature expectations of parents and possible wasted effort.

Future work will have to include testing the model in different ways. First, analyzing the model on a new dataset such as patient records obtained from June 2013 to the present. Second, once we finish operationalizing this model, we will collect provider feedback during daily rounds about their thoughts regarding a patient's discharge potential. We will then compare those results to the prediction of our model to determine if the providers or the machine-learning model is most accurate.

Conclusion

A supervised machine learning approach using a Random Forest classifier accurately predicts which patients will be discharged home from the NICU in the next 2-10 days. Running our model daily with the most recent progress note data will identify those patients who are close to being medically ready for discharge and may alert the clinical staff through indicators in the electronic medical record. This would allow for more timely discharge planning and has the potential to prevent delayed discharges due to non-medical reasons.

References

1. Bockli, K., et al., *Trends and challenges in United States neonatal intensive care units follow-up clinics.* J Perinatol, 2014. **34**(1): p. 71-74.
2. Challis, D., et al., *An examination of factors influencing delayed discharge of older people from hospital.* Int J Geriatr Psychiatry, 2014. **29**(2): p. 160-8.
3. Victor, C.R., et al., *Older patients and delayed discharge from hospital.* Health Soc Care Community, 2000. **8**(6): p. 443-452.
4. Szubski, C.R., et al., *Predicting discharge to a long-term acute care hospital after admission to an intensive care unit.* Am J Crit Care, 2014. **23**(4): p. e46-53.
5. Marcin, J.P., et al., *Long-stay patients in the pediatric intensive care unit.* Crit Care Med, 2001. **29**(3): p. 652-7.
6. Edwards, J.D., et al., *Chronic conditions among children admitted to U.S. pediatric intensive care units: their prevalence and impact on risk for mortality and prolonged length of stay*.* Crit Care Med, 2012. **40**(7): p. 2196-203.
7. Ruttimann, U.E. and M.M. Pollack, *Variability in duration of stay in pediatric intensive care units: a multiinstitutional study.* J Pediatr, 1996. **128**(1): p. 35-44.
8. Powell, P.J., et al., *When will my baby go home?* Arch Dis Child, 1992. **67**(10 Spec No): p. 1214-6.
9. Bannwart Dde, C., et al., *Prediction of length of hospital stay in neonatal units for very low birth weight infants.* J Perinatol, 1999. **19**(2): p. 92-6.
10. Lee, S.K., et al., *Variations in practice and outcomes in the Canadian NICU network: 1996-1997.* Pediatrics, 2000. **106**(5): p. 1070-9.
11. Lee, H.C., et al., *Accounting for variation in length of NICU stay for extremely low birth weight infants.* J Perinatol, 2013. **33**(11): p. 872-6.
12. Levin, S.R., et al., *Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders.* Crit Care Med, 2012. **40**(11): p. 3058-64.
13. http://scikit-learn.org/stable/index.html.
14. Koppel, R. and C.U. Lehmann, *Implications of an emerging EHR monoculture for hospitals and healthcare systems.* J Am Med Inform Assoc, 2014.
15. Kim, G.R. and C.U. Lehmann, *Pediatric aspects of inpatient health information technology systems.* Pediatrics, 2008. **122**(6): p. e1287-96.
16. Lehmann, C.U., *Pediatric aspects of inpatient health information technology systems.* Pediatrics, 2015. **135**(3): p. e756-68.

CHAPTER III

NATURAL LANGUAGE PROCESSING IMPROVES A DISCHARGE PREDICTION MODEL FOR THE NEONATAL ICU

Michael W. Temple[1], MD, Christoph U. Lehmann[1, 2], MD, Daniel Fabbri[1], PhD

Affiliations: [1]Department of Biomedical Informatics, [2]Department of Pediatrics Vanderbilt University, Nashville, TN.

Address correspondence to: Michael Temple, Department of Biomedical Informatics, Vanderbilt University School of Medicine, 2525 West End, Suite 1475, Nashville, TN 37203-8390, [michael.w.temple@vanderbilt.edu], 615-936-1068.

Short title: NLP Improves NICU Discharge Prediction Model.

Abbreviations: AUC – Area under the Curve, CART -- Classification And Regression Trees, DTD – Days to Discharge, GI – Gastrointestinal, LOS – Length of Stay, NICU – Neonatal Intensive Care Unit, NS – Neurosurgery, RF – Random Forest.

Key Words: Intensive Care Units, Neonatal; Area Under Curve; Patient Discharge; ROC Curve

Funding Source: National Library of Medicine Training Grant 5T15LM007450-13.

Financial Disclosure: Dr. Lehmann serves in a part-time role at the American Academy of Pediatrics. He also received royalties for the textbook Pediatric Informatics, and travel funds from the American Medical Informatics Association, the International Medical Informatics Association and the World Congress on Information Technology.
Dr. Fabbri has an equity interest in Maize Analytics, LLC. Dr. Temple has no financial disclosures.

Conflict of Interest: The authors have no conflicts of interest to disclose.

Abstract

Objectives
Discharging patients from the Neonatal Intensive Care Unit (NICU) can be delayed for non-medical reasons including the procurement of home medical equipment, parental education, and the need for children's services. We have previously created a model identify patients that will be medically ready for discharge in the next 2-10 days. In this study we use Natural Language Processing to improve that model and discern why that model performed poorly on some patients.

Materials and Methods
We retrospectively examined the text of the Assessment and Plan section from daily progress notes of 4,693 patient (103,206 patient-days) from the NICU of a large, academic children's hospital. A matrix was constructed using these words (single words and bigrams) and a supervised machine learning approach was used to determine the most important words differentiating poorly performing patients compared to well performing patients in our original discharge prediction model.

Results
NLP using a bag of words analysis revealed several cohorts that performed poorly in our original model. These included patients with surgical diagnoses, pulmonary hypertension, retinopathy of prematurity and psychosocial issues.

Discussion
The bag of words approach aided in cohort discovery and will allow for further refinement of our original discharge model prediction. Adequately identifying patients discharged home on g-tube feeds alone could improve the AUC of our original model by 0.02. Additionally, this approach identified social issues as causes for delayed discharge.

Conclusion
A bag of words analysis provides a method to improve and refine our NICU discharge prediction model and could potentially avoid over 900 (0.9%) hospital days.

Introduction

Approximately four million babies are born in the United States each year and approximately 11% of those are born prematurely.[1] The cost of caring for these infants can be substantial, with an estimated total annual cost of 26 billion dollars posing a significant financial burden for the health care system in general and hospitals specifically.[1] Discharging these patients as soon as they are medically ready is critical for controlling expenditures.

Delayed discharge of hospitalized patients who are medically ready for discharge is a common occurrence and often related to dependency and the need for post-discharge services.[2] Neonates discharge from the NICU are prime examples of patients with dependencies on parents and care-givers and who rely heavily on post-discharge services for medical follow-up, home medical equipment, and home nursing.[3] Parents of these fragile infants require a significant amount training and education regarding the special needs of their newborn, the use of medical equipment, and medication administration. These infants often require a number of services near discharge that may delay going home including hearing screens, repeat state screens, immunizations, car seat testing, and eye exams. Finally, infants at risk for abuse and neglect, for example with intra-uterine drug exposure, require consultation with Child Protective Services to ensure they are being discharged to a safe home environment.

We previously described a predictive model using a Random Forest to analyze 26 clinical features extracted from the NICU attending physician daily progress note.[3] The goal of that model was to identify patients who would be medically ready for discharge in the next 10, 7, 4, and 2 days so that the clinical staff would be aware and ready to address in advance the non-medical factors that often delay discharge of patients medically ready to go home.

This model performed well, achieving area under the curve (AUC) for the receiver operating characteristic (ROC) curve of 0.723, 0.754, 0.795, and 0.854 at 10, 7, 4 and 2 days until discharge, respectively. This model used structured and semi-structured data extracted from the attending physician progress note and it ignored the free text contained within the progress note. The goal of this current work is to use Natural Language Processing (NLP) to identify themes among poorly performing patients in our original model and to detect useful features missing from the original model. Using NLP along with expert domain knowledge should help us discover missing features to enable building a more accurate model for predicting when NICU patients are nearing discharge.

Related Work

NLP is a frequently used to analyze medical documentation in order to identify patient cohorts. Yang et al. describes a text mining approach for obesity detection and later expanded it to extract medication information.[4, 5] Jiang et al., in response to the 2010 Center of Informatics for Integrating Biology and the Bedside/Veterans Affairs challenge, examined different machine learning algorithms to identify clinical entities from discharge summaries.[6] Wright et al. used an NLP support vector machine to categorize free text notes in order to identify patients with diabetes.[7] In 2012, Cui et al. used discharge summaries to effectively extract information regarding epilepsy and seizure information.[8] Cosmin et al. describe an NLP system to identify ICU patients who were diagnosed with pneumonia at any point in their hospital stay.[9]

These studies demonstrated that NLP can be used to accurately identify patients belonging to certain cohorts. Typically when using NLP to evaluate the accuracy of a model, the results are compared to a known set of similar documents. This allows for the evaluation of precision, recall, and F-score. We propose to use NLP for cohort discovery. It is out hypothesis that NLP can assist us in refining our NICU prediction model and identify patient characteristics defined in the clinical note that may be missing in our original NICU discharge prediction model.

Methods

Patients and Setting

We conducted a retrospective study of all patients admitted to the NICU at a large academic medical center from June 2007 to May 2013.

Exclusion Criteria

Since this project was part of a larger study, the exclusion criteria were the same as the original study. All patients admitted to the NICU were considered for the study. Patients who were back-transferred to another facility or who died during the course of their NICU hospitalization were excluded from the analysis. Also excluded from the analysis were patients with any missing daily neonatology progress notes.

Data Collection and Extraction

A large database containing all of the daily progress notes written by neonatology attending physicians was made available to the investigators. The data from the progress notes were in a semi-structured text format that was extracted using regular expressions in Python (version 2.7.3) and SQL. In addition, these data were cross-referenced with the enterprise data warehouse in order to obtain basic patient information such as date of birth and ICD-9 codes used for billing during the hospitalization.

Feature Descriptions

Our original predictive model included the clinical features listed in Table 1.[3]

Table 1. Features used in the Predictive Model

Quantitative Features (Unit of Measure)	Qualitative Features (Unit of Measure)	Engineered Features (Unit of Measure)	Sub-Population Features
Weight (kg)	On Infused Medication (Y/N)	Number of Days Since Last A&B Event (days)	Premature (Y/N)
Birth Weight (kg)	On Caffeine (Y/N)	Number of Days Off Infused Medication (days)	Cardiac Surgery (Y/N)
Apnea and Bradycardia (A&B) Events (number)	On Ventilator (Y/N)	Number of Days Off Caffeine (days)	GI Surgery (Y/N)
Amount of Oral Feeds (ml)		Number of Days Off Ventilator (days)	Neurosurgery (Y/N)
Amount of Tube Feeds (ml)		Number of Days Off Oxygen (days)	
Percentage of Oral Feeds (%)		Number of Days Percent of Oral Feeds > 90% (days)	
Gestational Age (weeks)		Total Feeds (Oral + Tube Feeds) (ml)	
Gestational Age at Birth (weeks)		Ratio of Weight to Birth Weight	
Day of Life (days)		Amount of Oral Feeds / Weight (ml/kg/day)	
Oxygen (per liter)			

All of the clinical features listed in Table 1 were extracted using structured or semi-structured section of the progress note – not the Assessment and Plan. For the NLP evaluation, we used only the Assessment and Plan section of the daily progress note. This section tends to contain the most relevant clinical information.

The entire text of the Assessment and Plan section was extracted and tokenized using Python's natural language toolkit (version 3.0.1).[10] All of the stop words and numbers were removed. Additionally, words were converted to all lower case and only words with a length greater than or equal to three characters were considered in the corpus. This provided a simple "bag of words". Negation was not considered in this approach.

Matrix Generation

All of the extracted words were placed in a matrix (total number of words was 560). Each word was represented by a column. Each row represented one hospital day for a patient. Therefore, if the patient was in the hospital for 20 days, that patient occupied 20 rows of the matrix. If the word appeared in the Assessment and Plan section of the progress note on the day represented by that particular row, a '1' was assigned to the field representing the progress note and the patient. If the word was not present, a '0' was assigned.

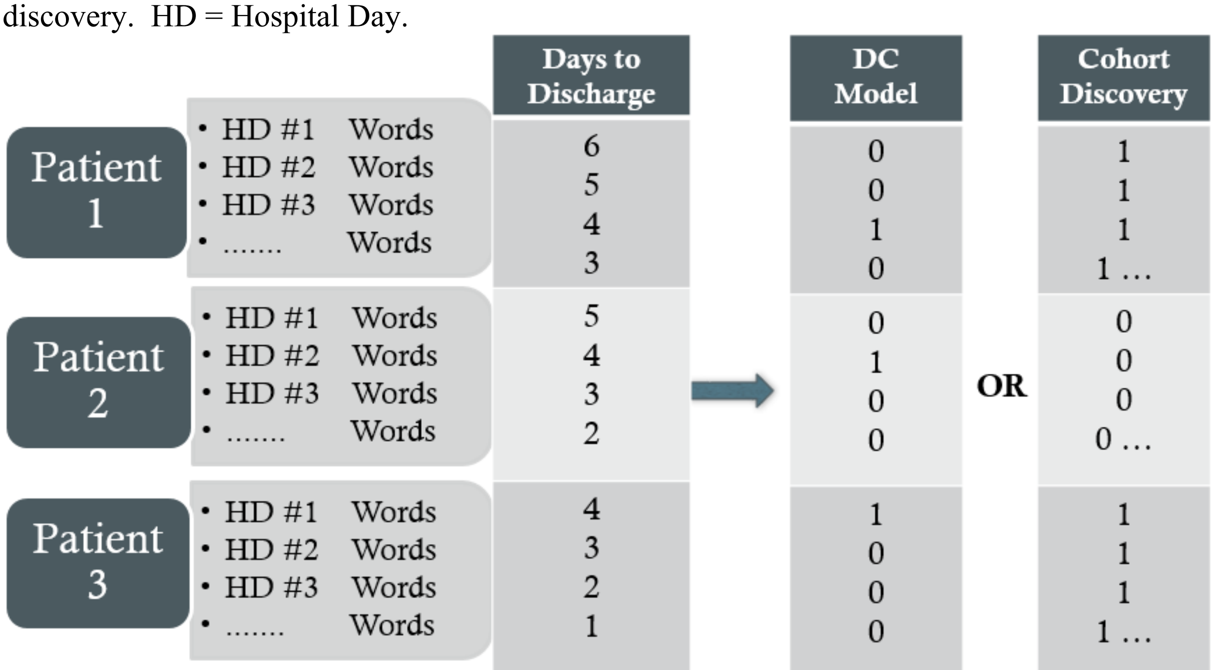
Model Vector Construction – Discharge Prediction

In addition to the columns for each word, there was also a column for days to discharge (DTD) . This column was used to build the dependent vector in the analysis (i.e. what we were trying to predict). For example, if we wanted to build a prediction model to determine which words were important if the patient was four days from discharge, then a '1' would be assigned in the DTD column when that patient was 4 days from discharge. For all other days for that patient, a '0' was assigned.

Model Vector Construction – Cohort Discovery

We were able to determine which patients had performed poorly or may have had a delayed discharge using the predicted probability of discharge from our discharge prediction original model. In this case, we assigned a '1' to the SP column for all the rows occupied by the group of poorly performing (or delayed discharge) patients and a '0' to the rows of patients that performed well. We then used this information to build a model to see if we could predict, using the bag of words from the Assessment and Plan, which patients would perform poorly or have a delayed discharge. See Figure 1.

Figure 1. Construction of matrix and model vector for predicting days to discharge or cohort discovery. HD = Hospital Day.

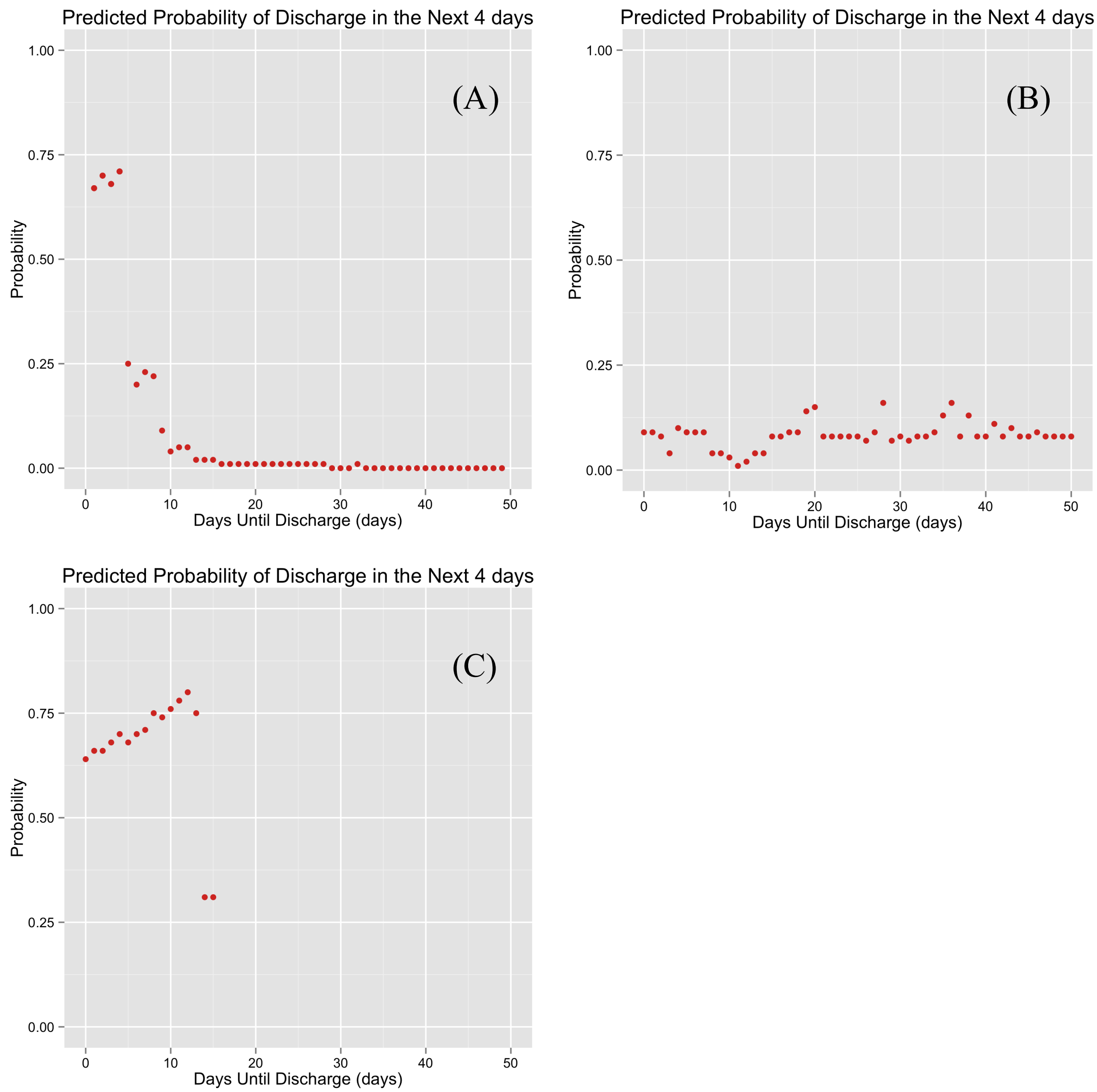


Data Analysis

A supervised machine learning approach using a Random Forest Classifier (RF) in Python's Sci-kit Learn module (version 0.15.2)[11] was used to analyze the data and build a predictive model. A RF constructs many binary decision trees that branch based on randomly chosen features. The RF in Sci-kit Learn uses an optimized Classification And Regression Trees (CART) algorithm for constructing binary trees using the features and thresholds (values) that yield the largest information gain at each node. The Sci-kit Learn package allows for the selection of either the gini impurity or entropy algorithms to determine feature importance. These algorithms performed similarly and we chose to use gini impurity because it is slightly more robust to misclassifications. We used the same Random Forest approach in our original model.

Models were trained using different combinations of DTD (2, 4, 7, 10 days) and different populations of poorly performing patients. Using our original prediction model, we were able to determine poorly performing patients by evaluating their predicted probability of discharge. For example, we ran our initial model predicting which patients were within 4 days of discharge from the NICU. We obtained the predicted probability (from 0 to 1) that our model assigned to each patient for each hospital day. If our model assigned a probability of 0.2 or less of discharge when the patient was actually 2 days from discharge, we then would consider this a poorly performing patient. Additionally, if our model assigned a probability of 0.5 or higher when the patient was 10 days or mode from discharge, these patients were considered delayed discharges. See Figure 2.

Figure 2. Graphs demonstrating the predicted probability of discharge from our original model. The patient is discharged when DTD = 0 (the left side of each graph). The right side of each graph are days early in the hospital stay. (A) Represents a patient classified as a "good performer". (B) Represents a "poor performer". (C) Represents a possible "delayed discharge".



Cross Validation

Each time a model was run, half of the patients (and all their associated daily rows) were randomized into a training set and the remaining patients were assigned to the testing set. Since the number of poorly performing patients in the SP was relatively small, halving the data provided both testing and training sets an adequate number of patients of interest. To achieve small enough standard deviations, the patients were randomized a total of five times for each model and the AUC for the ROC curve was obtained for the testing set. The reported AUC is the average of the five AUC's obtained after each round of randomization. Additionally, each time a model was run, the top 20 words used in the model were ranked in order of importance.

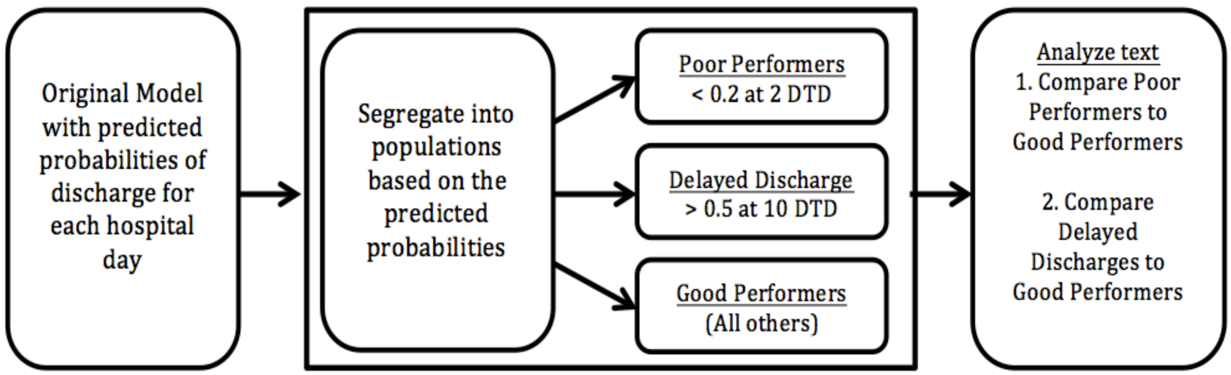
Model Generation

We ran the model for all patients to determine if a simple bag of words approach could outperform our original model for discharge prediction at 2, 4, 7, and 10 days from discharge. Additionally, we ran the model comparing patients that performed well in our original model to those that performed poorly in our original model. Finally, the most important words contained in the Assessment and Plan section of the daily progress note at 2, 4, 7, and 10 days to discharge were determined as well as the most important words differentiating poorly performing patients to those that performed well in our original model. We determined the poor performers from the original model by the following steps (See Figure 3):

1. We ran the original model predicting which patients would be ready for discharge in the next 4 days.
2. The prediction model outputted a probability for each row in the matrix (a row consisted of a single hospital day for a single patient).
3. We then obtained the patient identifier of those patients that the model assigned a probability of 0.2 or less for that patient being discharged in the next two days (or a probability of 0.5 or greater at days to discharge of 10 or more).

4. These patients were then used as the classifier for the Random Forest prediction. The words that were most important for the prediction were then returned. We used single words as well as bigrams.

Figure 3. Workflow diagram demonstrating process for cohort discovery.



IRB Approval

The Institutional Review Board of Vanderbilt University approved this study.

Results

The initial database consisted of 6,302 patients admitted to the NICU between June 2007 and May 2013. There were 256 deaths during this time period. A total of 1,154 patients were excluded because the database did not contain physician progress notes for every day of their hospital course. There were 199 patients back-transferred to other NICU's in the region. The final matrix consisted of 4,693 unique patients accounting for 103,206 hospital days with a mean LOS of 30 days.

Bag of Words for Discharge Prediction

Table 2 shows the results of the original model only, bag of words (BOW) only, and the combined approach using only words from the Assessment and Plan with regards to discharge prediction.

Table 2. Comparing discharge prediction models among the original model, BOW model and the combination of the two models. BOW = bag of words.

Days Until Discharge (days)	Original Model (AUC)	BOW Model (AUC)	Combined Original and BOW (AUC)
10	0.723	0.569	0.633
7	0.754	0.589	0.677
4	0.795	0.654	0.752
2	0.854	0.743	0.837

Table 3 shows the top 15 most important bigrams for predicting discharge at 2, 4, 7, and 10 days until discharge.

Table 3. The top 15 most important (listed in order) bigrams for each of the days to discharge listed

Days Until Discharge (days)	Most important Bigrams
10	continue monitor, today continue, pcv retic, enteral feeds, day continue, total fluids, prior discharge, feeds day, weight gain, continue follow, past hrs, full feeds, updated bedside, wean today, room air
7	continue monitor, weight gain, prior discharge, today continue, pcv retic, full feeds, enteral feeds, feeds day, next week, day continue, past hours, amp gent, may need, continue follow, past hrs
4	prior discharge, continue monitor, weight gain, pcv retic, today continue, feeds day, past hrs, day continue, cbc crp, amp gent, room air, follow clinically, past hours, discharge home, continue follow
2	weight gain, prior discharge, continue monitor, full feeds, pcv retic, hearing screen, room air, amp gent, fen lib, repeat echo, cbc crp, continue follow, today continue, last hours, follow clinically.

Bag of Words for Cohort Discovery – Probability less than 0.2 at 2 or less DTD

We extracted the most important words as determined by the bag of words model when comparing patients who performed well in our original model to those that performed poorly in our original model.

Table 4 shows the most significant words differentiating well performing from poorly performing patients with a probability of 0.2 or less to be discharged in the next two days. The words are listed in order of importance and a few words have been excluded because of inability to determine the context (for example, “continue monitor”, and “per protocol”).

Table 4. The most important single words and bigram differentiating poorly performing patients (probability of less than 0.2 at 2 or less days until discharge) from well performing patients in our original model. Listed in order of importance.

Single Words	Bigrams
fistula, ent, tube, esophageal, atresia, nissen, vfss, breech, psychosocial, uti, gtube, aspiration, hus, reflux, vcug	status post, esophageal atresia, repeat echo, pulmonary hypertension, enteral feeds, lung disease, goal sats, urine culture, infectious disease, drug screen, plus disease, stage zone, room air

Bag of Words for Cohort Discovery – Probability more than 0.5 at 10 or more DTD

Table 5 lists the most significant words differentiating poorly performing patients with a probability of 0.5 or higher at 10 or more days until discharge.

Table 5. The most important single words and bigram differentiating poorly performing patients (probability of more than 0.5 at 10 or more days until discharge) from well performing patients in our original model. Listed in order of importance.

Single Words	Bigrams
hep, social, weight, daily, restarted, signs, direct, endocrine, positive, drug, mother, birth, dcs, congenital, syndrome, continue, prematurity	social work, work breathing, low birth, birth weight, initial cbc, clinical signs, room air, dcs involved, possible sepsis, prior discharge, infectious disease, monitor respiratory, continue monitor, hearing screen, newborn screen, meconium drug, drug screen

Discussion

Bag of Words for Discharge Prediction

The bag of words approach, not surprisingly, performed poorly with regards to discharge prediction. This may be explained by the fact that only a very small part of the progress note (the Assessment and Plan section) was used as the corpus. If only the bag of words approach were to be used as the sole prediction model, then the entire daily progress note would have been used. Second, because our original model contained quantitative clinical data, we excluded any numerical values from out NLP analysis.

Bag of Words for Cohort Discovery – Probability less than 0.2 at 2 or less DTD

Using a bag of words model for cohort discovery identified characteristics for some patients that are not performing well in our original model (See Table 4).

First, our original model is not performing well on some surgical patients. The top two most important bigrams are "status post" and "esophageal atresia". Additionally, four of the most important single words are "fistula", "esophageal", "atresia", and "nissen". All of these words would be found in patients who have a gastrointestinal abnormality requiring surgery or have had a surgical repair already performed. Feeding difficulties and subsequent increased length of stay have been described in this population.[12] Also, patients who have had a "nissen" procedure likely needed the procedure because of reflux with aspiration pneumonia. The words "aspiration", "reflux", "gtube" and "vfss" (swallow study) are likely related to this GI surgery. Finally, one of the most important single words is "ent". Neonates can have congenital anomalies of their ear, nose or throat requiring surgical correction; therefore, capturing these patients in our model could help improve it.

Another interesting combination of words for cohort discovery is “psychosocial” and “drug screen”. The importance of these words would seem to indicate that our model is not performing well on patients who may have had intrauterine drug exposure or whose parents may have had psychosocial issues.

Our model also appears to perform poorly on patients who have a history of “pulmonary hypertension”. These patients tend to be very sick early in their hospital stay and may require extra-corporeal membrane oxygenation (ECMO). While these patients have significantly improved clinical status when they are two days from discharge, it appears that our model is not correctly capturing the improved clinical status of these patients.

Finally, the two bigrams “plus disease” and “stage zone” are references to retinopathy of prematurity. Premature infants with retinopathy of prematurity (ROP) need to have an eye exam performed by an ophthalmologist near the time of their discharge. The presence of these words in the Assessment and Plan could be referencing the results of this last exam before discharge or the need to schedule an examination prior to discharge.

Bag of Words for Cohort Discovery – Probability more than 0.5 at 10 or more DTD

Using a bag of words approach on these patients helped identify possible reasons for patients that may have their discharges delayed (See Table 5). First, social factors appear to be an issue. Words such as “social”, “drug”, and “dcs” (Department of Children’s Services) indicate social and/or custody issues may be causing discharge delays in patients who are medically ready for discharge. This is further supported by the bigrams “social work”, “dcs involved”, “meconium drug”, and “drug screen”.

In addition to our original model predicting a greater than 0.5 probability of discharge for these patients, the bag of words also supports their readiness for discharge. Words from Table 3 (important words for discharge prediction) such as “prior discharge”, “continue monitor”, “room air”, “hearing screen” also appear in table 5 – the list of important words for patients who may be ready for discharge, but are delayed. In our data set, there were 904 hospital days (198 patients) that met these probability criteria. Both the original model and NLP analysis would suggest that potentially 904 (0.9%) hospital days could have been avoided in these patients who likely had delays in their discharge.

Further Evaluation

The bag of words approach certainly identified patient characteristics that were not present in our original model mainly pertaining to specific diagnoses that lead to feeding problems or need for prolonged monitoring like ROP. Using this knowledge in our model we will be able to add other features that will aid to capture and improve the predictive accuracy of these poorly performing patients. For example, our model could identify patients that have had a social work consult performed. We could also use ICD-9 codes to capture patients who have esophageal atresia, pulmonary hypertension, or retinopathy of prematurity.

In our original model, important predictive factors centered around feeding – in particular oral feeding. If the infant was consistently consuming a large part of their feeds orally, then they were nearing discharge. This NLP analysis would indicate that our model is not performing well on patients who go home on g-tube feedings. Therefore, we performed the following test to determine the impact on our model if we correctly classified those patients being discharged on g-tube feeds:

1. We used the NLP bag of words approach and identified all patients who had the words “gtube” or “g-tube” in Assessment and Plan of their progress note.
2. We then used these patient identifiers in our original model.
3. We ran our original model as normal, except when the model was creating the output (prediction) vector, if the patient was in the “g-tube” cohort, we ensured that the output vector contained a ‘1’ and not a ‘0’ (predicting the patient is near discharge).

The result of this manipulation of the output vector is shown in Table 6.

Table 6. The improvement our original model would show if we were able to correctly capture and classify all patients who were discharged home on g-tube feeds.

Days Until Discharge (days)	**Original Model (AUC)**	**Correctly classified g-tube patients (AUC) (difference)**
10	0.723	0.741 (+ 0.018)
7	0.754	0.775 (+ 0.021)
4	0.795	0.817 (+ 0.022)
2	0.854	0.863 (+ 0.009)

Table 6 demonstrates that correctly classifying patients who are discharged home on g-tube feeds improves the accuracy of our predictive model.

Limitations and Next Steps

One limitation of this study is that we only used the Assessment and Plan section of the attending physician progress note in the bag of words model. It is likely that more information from the use of the entire progress note would be benefit the accuracy of our predictive model.

Another limitation is that even though NLP identified cohorts that do not perform well in our original model, it may be difficult to find a way to integrate those cohorts in our original model. For example, some patients who are discharge home on g-tube feeds may actually look different clinically. Some patients may be able to take a portion of their feedings orally while others will be reliant on continuous g-tube feedings.

A final limitation with an NLP analysis performed is that not all patients may be correctly classified. For example, while we identified a significant word as “vfss”, there may be other patients in whom “swallow study” is actually written out in the assessment and plan. Capturing all the ways in which medical professionals abbreviate is a difficult task and can cause some patients to be misclassified.

The next steps in the refinement of our NICU discharge prediction model will be to use these cohorts discovered through our bag of words analysis and modify our original prediction model to include features related to these cohorts. For example, we could use ICD-9 codes to capture patients with pulmonary hypertension and retinopathy of prematurity to determine if there are other features that can be used to more accurately classify these patients.

Conclusions

An NLP analysis using a simple bag of words approach can be effectively used to discover under-performing cohorts and delayed discharges in a NICU discharge prediction model. Correctly classifying these cohorts can then be used to improve the predictive accuracy of the model and, in the case of the delayed discharges, avoid over 900 hospital days.

References

1. Bockli, K., et al., *Trends and challenges in United States neonatal intensive care units follow-up clinics.* J Perinatol, 2014. **34**(1): p. 71-74.
2. Challis, D., et al., *An examination of factors influencing delayed discharge of older people from hospital.* Int J Geriatr Psychiatry, 2014. **29**(2): p. 160-8.
3. Temple, M.W., Lehmann, C.U., Fabbri, D., *Using Daily Progress Note Data to Predict Discharge Date from the Neonatal Intensive Care Unit.* Accepted by *Pediatrics.* Publication Pending.
4. Yang, H., et al., *A text mining approach to the prediction of disease status from clinical discharge summaries.* J Am Med Inform Assoc, 2009. **16**(4): p. 596-600.
5. Yang, H., *Automatic extraction of medication information from medical discharge summaries.* J Am Med Inform Assoc, 2010. **17**(5): p. 545-8.
6. Jiang, M., et al., *A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries.* J Am Med Inform Assoc, 2011. **18**(5): p. 601-6.
7. Wright, A., et al., *Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions.* J Am Med Inform Assoc, 2013. **20**(5): p. 887-90.
8. Cui, L., et al., *EpiDEA: extracting structured epilepsy and seizure information from patient discharge summaries for cohort identification.* AMIA Annu Symp Proc, 2012. **2012**: p. 1191-200.
9. Bejan, C.A., et al., *On-time clinical phenotype prediction based on narrative reports.* AMIA Annu Symp Proc, 2013. **2013**: p. 103-10.
10. http://www.nltk.org.
11. http://scikit-learn.org/stable/index.html.
12. Wang, J., et al., *Prolonged feeding difficulties after surgical correction of intestinal atresia: a 13-year experience.* J Pediatr Surg, 2014. **49**(11): p. 1593-7.

CHAPTER IV

SUMMARY

Predicting when a patient will be discharged from the NICU is a challenging task. There is great variability in conditions seen in the NICU and many of these patients have a prolonged length of stay. Additionally, planning for the discharge of these complex patients is a difficult and time-consuming task. This complexity can delay discharges from the NICU in patients that are otherwise medically ready for home. The focus of this project was to identify in advance those patients who are nearing discharge in order to provide the clinical staff the needed time to adequately prepare the infant and care givers for this important transition.

Specific Aim #1 was addressed in the first manuscript. This Random Forest model using clinical data from the attending physician progress note proved to be accurate in predicting which patients are nearing discharge. This should allow the clinical staff adequate notice of the impending discharge and give them enough lead time to prepare the infant and parents for discharge.

Specific Aim #2 was also addressed in the first manuscript. The predictive model was able to identify which features were the most important for predictive accuracy. The flexibility of this model allowed for the construction of a simple decision tree using only 2 features that was nearly as accurate as the model including all the features extracted. This simple decision tree could easily be used at the bedside as a “rule-of thumb” by the clinical team to get a general sense about the infant’s readiness for discharge.

Specific Aim #3 was the focus of the second manuscript. Using a bag of words on a portion of the progress note allowed for the identification of several cohorts that did not perform well in the original model. This type of NLP analysis could certainly provide a framework for cohort discovery and refinement of the predictive model.

APPENDIX I

ICD code	Description	Category
746.01	atresia of pulmonary valve, congenital	Cardiac
747.49	other anomalies of great veins	Cardiac
428	congestive heart failure, unspecified	Cardiac
428.2	systolic heart failure, unspecified	Cardiac
429	myocarditis, unspecified	Cardiac
429.3	cardiomegaly	Cardiac
745.1	complete transposition of great vessels	Cardiac
745.1	complete transposition of great vessels	Cardiac
745.11	double outlet right ventricle	Cardiac
745.2	tetralogy of fallot	Cardiac
427.89	other specified cardiac dysrhythmias, other	Cardiac
745.6	endocardial cushion defect, unspecified type	Cardiac
427.42	ventricular flutter	Cardiac
746.02	stenosis of pulmonary valve, congenital	Cardiac
746.09	other congenital anomalies of pulmonary valve	Cardiac
746.3	congenital stenosis of aortic valve	Cardiac
746.4	congenital insufficiency of aortic valve	Cardiac
746.87	malposition of heart and cardiac apex	Cardiac
746.89	other specified congenital anomalies of heart	Cardiac
746.9	unspecified congenital anomaly of heart	Cardiac
747.1	coarctation of aorta (preductal) (postductal)	Cardiac
747.21	congenital anomalies of aortic arch	Cardiac
747.3	congenital anomalies of pulmonary artery	Cardiac
745.4	ventricular septal defect	Cardiac
424.9	endocarditis, valve unspecified, unspecified cause	Cardiac
396.3	mitral valve insufficiency and aortic valve insufficiency	Cardiac
397	diseases of tricuspid valve	Cardiac
420.9	acute pericarditis, unspecified	Cardiac
420.99	other acute pericarditis	Cardiac
421	acute and subacute bacterial endocarditis	Cardiac
422.91	idiopathic myocarditis	Cardiac
423.3	cardiac tamponade	Cardiac
424	mitral valve disorders	Cardiac
424.1	aortic valve disorders	Cardiac
427.9	cardiac dysrhythmia, unspecified	Cardiac
424.3	pulmonary valve disorders	Cardiac
745.3	common ventricle	Cardiac
425.1	hypertrophic cardiomyopathy	Cardiac
425.3	endocardial fibroelastosis	Cardiac

425.4	other primary cardiomyopathies	Cardiac
425.8	cardiomyopathy in other diseases classified elsewhere	Cardiac
426	atrioventricular block, complete	Cardiac
426.1	atrioventricular block, unspecified	Cardiac
426.11	first degree atrioventricular block	Cardiac
426.12	mobitz (type) ii atrioventricular block	Cardiac
426.13	other second degree atrioventricular block	Cardiac
427.41	ventricular fibrillation	Cardiac
424.2	tricuspid valve disorders, specified as nonrheumatic	Cardiac
V15.1	personal history of surgery to heart and great vessels, presenting hazards to health	Cardiac
794.3	unspecified nonspecific abnormal function study of cardiovascular system	Cardiac
794.39	other nonspecific abnormal function study of cardiovascular system	Cardiac
997.1	cardiac complications, not elsewhere classified	Cardiac
745.12	corrected transposition of great vessels	Cardiac
997.79	vascular complications of other vessels	Cardiac
777.1	meconium obstruction in fetus or newborn	GI Surgery
530.3	stricture and stenosis of esophagus	GI Surgery
530.4	perforation of esophagus	GI Surgery
530.6	diverticulum of esophagus, acquired	GI Surgery
777.5	necrotizing enterocolitis in newborn, unspecified	GI Surgery
530.89	other specified disorders of the esophagus	GI Surgery
777.51	stage i necrotizing enterocolitis in newborn	GI Surgery
553.1	umbilical hernia without mention of obstruction or gangrene	GI Surgery
557.9	unspecified vascular insufficiency of intestine	GI Surgery
560.2	volvulus	GI Surgery
560.81	intestinal or peritoneal adhesions with obstruction (postoperative) (postinfection)	GI Surgery
560.89	other specified intestinal obstruction, other	GI Surgery
569.83	perforation of intestine	GI Surgery
569.69	other colostomy and enterostomy complication	GI Surgery
530.84	tracheoesophageal fistula	GI Surgery
756.79	other congenital anomalies of abdominal wall	GI Surgery
751.3	hirschsprung's disease and other congenital functional disorders of colon	GI Surgery
751.2	congenital atresia and stenosis of large intestine, rectum, and anal canal	GI Surgery
751.1	congenital atresia and stenosis of small intestine	GI Surgery
750.4	other specified congenital anomalies of esophagus	GI Surgery
V55.2	attention to ileostomy	GI Surgery
756.72	congenital anomalies of abdominal wall, omphalocele	GI Surgery

V55.4	attention to other artificial opening of digestive tract	GI Surgery
756.73	congenital anomalies of abdominal wall, gastroschisis	GI Surgery
560.9	unspecified intestinal obstruction	GI Surgery
777.53	stage iii necrotizing enterocolitis in newborn	GI Surgery
777.52	stage ii necrotizing enterocolitis in newborn	GI Surgery
777.5	necrotizing enterocolitis in newborn, unspecified	GI Surgery
V55.1	attention to gastrostomy	GI Surgery
V44.1	gastrostomy status	GI Surgery
536.49	other gastrostomy complications	GI Surgery
536.42	mechanical complication of gastrostomy	GI Surgery
536.41	infection of gastrostomy	GI Surgery
742.9	unspecified congenital anomaly of brain, spinal cord, and nervous system	Neurosurgery
741	spina bifida, unspecified region, with hydrocephalus	Neurosurgery
331.3	other cerebral degenerations, communicating hydrocephalus	Neurosurgery
331.4	other cerebral degenerations, obstructive hydrocephalus	Neurosurgery
742.4	other specified congenital anomalies of brain	Neurosurgery
742.3	congenital hydrocephalus	Neurosurgery
741.9	spina bifida, unspecified region, without mention of hydrocephalus	Neurosurgery
741.02	spina bifida, dorsal (thoracic) region, with hydrocephalus	Neurosurgery
741.03	spina bifida, lumbar region, with hydrocephalus	Neurosurgery
742.1	microcephalus	Neurosurgery
741.93	spina bifida, lumbar region, without mention of hydrocephalus	Neurosurgery
552.3	diaphragmatic hernia with obstruction	PPH/ECMO
756.6	congenital anomalies of diaphragm	PPH/ECMO
747.83	congenital anomaly, persistent fetal circulation	PPH/ECMO
416	primary pulmonary hypertension	PPH/ECMO
763.84	meconium passage during delivery affecting fetus or newborn	PPH/ECMO
764.94	unspecified fetal growth retardation, 1000-1249 grams	Premature
765.01	disorders relating to extreme immaturity of infant, less than 500 grams	Premature
362.24	retinopathy of prematurity, stage 2	Premature
779.7	periventricular leukomalacia	Premature
764.95	unspecified fetal growth retardation, 1250-1499 grams	Premature
765	disorders relating to extreme immaturity of infant, weight unspecified	Premature
764.92	unspecified fetal growth retardation, 500-749 grams	Premature
772.13	intraventricular hemorrhage of fetus or newborn, grade iii	Premature

765.02	disorders relating to extreme immaturity of infant, 500-749 grams	Premature
362.25	retinopathy of prematurity, stage 3	Premature
772.12	intraventricular hemorrhage of fetus or newborn, grade ii	Premature
362.23	retinopathy of prematurity, stage 1	Premature
362.21	retrolental fibroplasia	Premature
362.2	retinopathy of prematurity, unspecified	Premature
362.27	retinopathy of prematurity, stage 5	Premature
765.28	disorders related to weeks of gestation completed, 35-36 weeks	Premature
765.17	disorders relating to other preterm infants, 1750-1999 grams	Premature
765.16	disorders relating to other preterm infants, 1500-1749 grams	Premature
765.15	disorders relating to other preterm infants, 1250-1499 grams	Premature
765.18	disorders relating to other preterm infants, 2000-2499 grams	Premature
765.22	disorders related to weeks of gestation completed, 24 weeks	Premature
765.24	disorders related to weeks of gestation completed, 27-28 weeks	Premature
765.25	disorders related to weeks of gestation completed, 29-30 weeks	Premature
776.6	anemia of prematurity	Premature
765.27	disorders realted to weeks of gestation completed, 33-34 weeks	Premature
765.03	disorders relating to extreme immaturity of infant, 750-999 grams	Premature
769	respiratory distress syndrome in newborn	Premature
770.7	chronic respiratory disease arising in the perinatal period	Premature
772.1	intraventricular hemorrhage of fetus or newborn, unspecified grade	Premature
772.11	intraventricular hemorrhage of fetus or newborn, grade i	Premature
772.14	intraventricular hemorrhage of fetus or newborn, grade iv	Premature
765.14	disorders relating to other preterm infants, 1000-1249 grams	Premature
765.13	disorders relating to other preterm infants, 750-999 grams	Premature
765.1	disorders relating to other preterm infants, weight	Premature

	unspecified	
765.26	disorders related to weeks of gestation completed, 31-32 weeks	Premature