

Predicting Discharge Dates From the NICU Using Progress Note Data

Michael W. Temple, MD^a, Christoph U. Lehmann, MD^{a,b}, Daniel Fabbri, PhD^a

abstract

BACKGROUND AND OBJECTIVES: Discharging patients from the NICU may be delayed for nonmedical reasons including the need for medical equipment, parental education, and children's services. We describe a method to predict which patients will be medically ready for discharge in the next 2 to 10 days, providing lead time to address nonmedical reasons for delayed discharge.

METHODS: A retrospective study examined 26 features (17 extracted, 9 engineered) from daily progress notes of 4693 patients (103 206 patient-days) from the NICU of a large, academic children's hospital. These data were used to develop a supervised machine learning problem to predict days to discharge (DTD). Random forest classifiers were trained by using examined features and International Classification of Diseases, Ninth Revision–based subpopulations to determine the most important features.

RESULTS: Three of the 4 subpopulations (premature, cardiac, gastrointestinal surgery) and all patients combined performed similarly at 2, 4, 7, and 10 DTD with area under the curve (AUC) ranging from 0.854 to 0.865 at 2 DTD and 0.723 to 0.729 at 10 DTD. Patients undergoing neurosurgery 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. By using DTD equal to 4 and 2 features (oral percentage of feedings and weight), we constructed a model with an AUC of 0.843.

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

Departments of ^aBiomedical Informatics, and ^bPediatrics, Vanderbilt University, Nashville, Tennessee

Dr Temple drafted the manuscript, contributed to the data collection, analysis, and model development, reviewed and revised the manuscript, and prepared it for publication; Dr Lehmann assisted with the data collection, aided in the selection of relevant clinical features for the model, categorized ICD-9 codes for grouping patients into distinct populations, and reviewed and revised the manuscript; Dr Fabbri assisted with the data collection, analysis, and model development, contributed to the machine learning and statistical analysis of the data, and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

www.pediatrics.org/cgi/doi/10.1542/peds.2015-0456

DOI: 10.1542/peds.2015-0456

Accepted for publication May 19, 2015

Address correspondence to Michael W. Temple, MD, Department of Biomedical Informatics, Vanderbilt University School of Medicine, 2525 West End, Suite 1475, Nashville, TN 37203-8390. E-mail: michael.w.temple@vanderbilt.edu, mtemple1@me.com

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2015 by the American Academy of Pediatrics

WHAT'S KNOWN ON THIS SUBJECT: Discharge from the NICU requires coordination and may be delayed for nonmedical 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 using real-time patient data from the daily neonatology progress note to predict when patients will be medically ready for discharge.

Approximately 4 million babies are born every year in the United States, and about 11% (~440 000) of them are born prematurely.¹ Caring for infants in the NICU poses a significant financial burden to the health care system, with an estimated total cost of \$26 billion.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 postdischarge services.² In older adults, difficulties in coordinating postdischarge 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 postdischarge needs such as primary care, specialists, physical and speech therapy, neonatal follow-up appointments, home equipment services, and home nursing. In cases of intrauterine drug exposure, discharge is often dependent on Child Protective Services approval. Parents have to demonstrate their ability to operate medical equipment, administer home medication, and 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 examinations. All 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 patients who are close to discharge from

a medical perspective so staff can be alerted to impending discharges. Doing so will allow the nonmedical factors to be addressed in advance to ensure that the patient's discharge is not 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} Although 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 patients in the NICU are likely to be discharged from the hospital in the next 2 to 10 days. Our method predicts the upcoming discharge date, not the LOS from time of admission.

To prevent delayed discharge, 3 questions will be answered. First, can the discharge date for a patient in the NICU be accurately predicted? Second, what combinations of clinical data improve predictive accuracy? Third, are there simple, "rule-of-thumb" factors that are responsible for a substantial fraction of the prediction accuracy?

Because of the potential impact on cost savings, predicting the LOS for patients in the NICU has been well studied. Most of the following prediction methods were performed at or near the time of admission. Powell et al⁸ found gestational age, low birth weight, and respiratory difficulties to be most predictive of LOS. Bannwart et al⁹ developed 2 models to predict the LOS for patients in the NICU. The first model considered only risk factors present in the first 3 days of life, whereas 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. Studying the Canadian NICU Network, Lee et al¹⁰ found that "significant variation in NICU practices and outcomes was observed despite Canada's universal health insurance system." Using data from the California Perinatal Quality Care Collaborative, Lee et al¹¹ reported "wide variance in LOS by birth weight, gestational age, and other factors."

In 2012, Levin et al¹² described a real-time model to forecast LOS in a PICU by using physician orders from a provider order entry system. 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 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 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 daily progress notes written by neonatology attending physicians was made available to the investigators. The data from the progress notes were in a semistructured text format that was extracted through regular expressions in Python version 2.7.3 (Python Software Foundation, Beaverton, OR) and SQL. In addition, these data were cross-referenced with the enterprise data warehouse to obtain basic patient information such as date of birth and International Classification of Diseases, Ninth Revision (ICD-9) codes used for billing during the hospitalization.

Feature Descriptions

The clinical features used in our model fell into 4 main categories: quantitative, qualitative, engineered, and derived subpopulations. Thirteen features were obtained directly from

data contained in 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&Bs) in the last 24 hours. The engineered feature from these data was the number of days since the last A&B.

Additionally, a neonatologist (C.U.L.) reviewed 138 of the most frequently occurring ICD-9 codes in the NICU patient population to categorize patients into 4 subpopulations: prematurity, cardiac disease, gastrointestinal (GI) surgical disease, and neurosurgical (NS) disease (see the Appendix for a list of ICD-9 codes and categories). A single patient could belong to 1, many, or none of the subpopulations. Table 1 contains a list of all features used in the model.

Matrix Generation

All extracted data, subpopulation categories, engineered features, and days to discharge (DTD) were inserted

into a matrix. Each row represented data for 1 hospital day for a specific patient. If a row contained missing data in any field, the entire row was excluded from the final matrix.

Because 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 (Fig 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. An 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 by using the input features and values that yield the largest information gain at each node. The Sci-kit Learn package allows 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 with different combinations of subpopulations (all patients, premature, cardiac, GI surgical, and NS), DTD (2, 4, 7, and 10 days), and number of features (any combination of features from 2 to all 26).

TABLE 1 Features Used in the Predictive Model

Quantitative Features (Unit of Measure)	Qualitative Features (Unit of Measure)	Engineered Features (Unit of Measure)	Subpopulation Features
Wt (kg)	On infused medication (Y/N)	Number of days since last A&B (d)	Premature (Y/N)
Birth wt (kg)	On caffeine (Y/N)	Number of days off infused medication (d)	Cardiac surgery (Y/N)
A&Bs (no.)	On ventilator (Y/N)	Number of days off caffeine (d)	GI surgery (Y/N)
Amount of oral feeds (mL)	—	Number of days off ventilator (d)	Neurosurgery (Y/N)
Amount of tube feeds (mL)	—	Number of days off oxygen (d)	—
Percentage of oral feeds (%)	—	Number of days percentage of oral feeds >90% (d)	—
Gestational age (wk)	—	Total feeds (oral + tube feeds) (mL)	—
Gestational age at birth (wk)	—	Ratio of wt to birth wt	—
Day of life (d)	—	Amount of oral feeds/wt (mL/kg/day)	—
Oxygen (L)	—	—	—

—, not applicable.

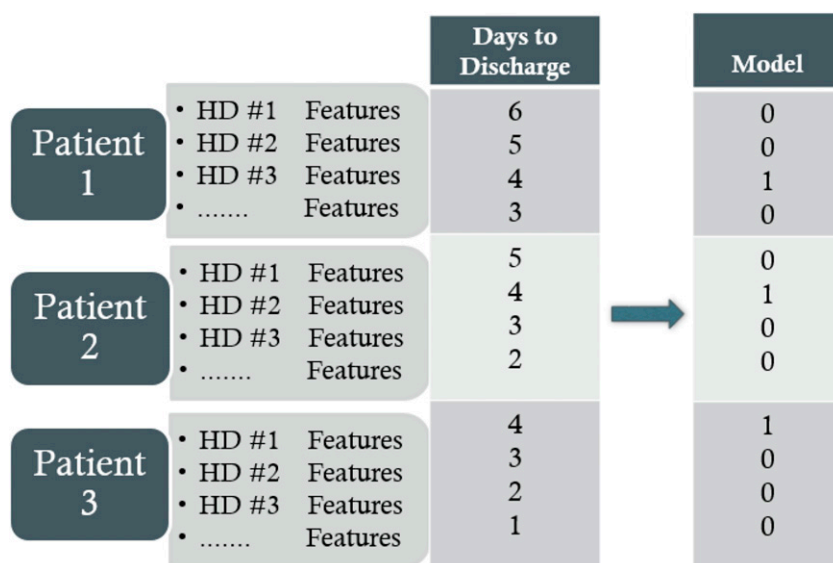


FIGURE 1
Example data matrix construction showing an attempt to model 4 DTD. HD, hospital day.

Training Vector

To train our model, we converted the DTD 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 4 days from discharge, all values in the model where the DTD was not equal to 4 were set to 0. The rows in which the number of DTD was 4 were set to 1 (Fig 1). This same process was followed for 2, 7, and 10 DTD.

Cross-Validation

Each time a model was run, half of the patients (and all their associated daily rows) were randomly assigned to a training set, and the other half were assigned to the testing set. Because 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 randomly assigned 5 times for each model and the area under the curve (AUC) for the receiver operating characteristic curve was obtained for the testing set. The reported AUC is the average of the 5 AUCs 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 subpopulation to determine how well the model performed, to choose the most important features for each group, and to determine whether different features had a greater impact on certain patient populations. Finally, the most important features at 2, 4, 7, and 10 DTD were evaluated to determine whether the most important features changed as a patient was getting closer to discharge.

Institutional Review Board Approval

The Institutional Review Board of Vanderbilt University approved this study.

RESULTS

The initial database consisted of 6302 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 1154 (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 NICUs in the region. The final matrix consisted of 4693 (74%) unique patients, accounting for 103 206 (89%) hospital days with a mean LOS of 30 days. A total of 3689 (79%) patients were categorized into ≥ 1 subpopulations based on ICD-9 codes; the other 1004 (21%) patients did not have an ICD-9 code that matched our criteria (Fig 2).

The average AUC for the model using all 26 features for all patients and each patient subpopulation is shown in Fig 3. Three of the 4 subpopulations (premature, cardiac, GI surgery) and all patients combined performed very similarly at 2, 4, 7, and 10 DTD, with AUCs ranging from 0.854 to 0.865 at 2 DTD and 0.723 to 0.729 at 10 DTD. The NS subpopulation performed worse on every DTD measure, scoring 0.749 at 2 DTD and 0.614 at 10 DTD (Fig 3). Using fivefold cross-validation provided a sufficiently narrow SD range for AUCs of ~ 0.005 to 0.01.

The 9 most predictive features for each subpopulation were very similar, and their plots are shown in Fig 4. In each subpopulation, the combination of all features performed better than any single feature alone. Once again, the poorest-performing subpopulation included the NS patients.

In addition to analyzing the most important features for each subpopulation, we 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 patients moved closer to discharge.

DISCUSSION

We were able to use data from daily progress notes to predict impending discharge from the NICU accurately. Our model improved as more clinical

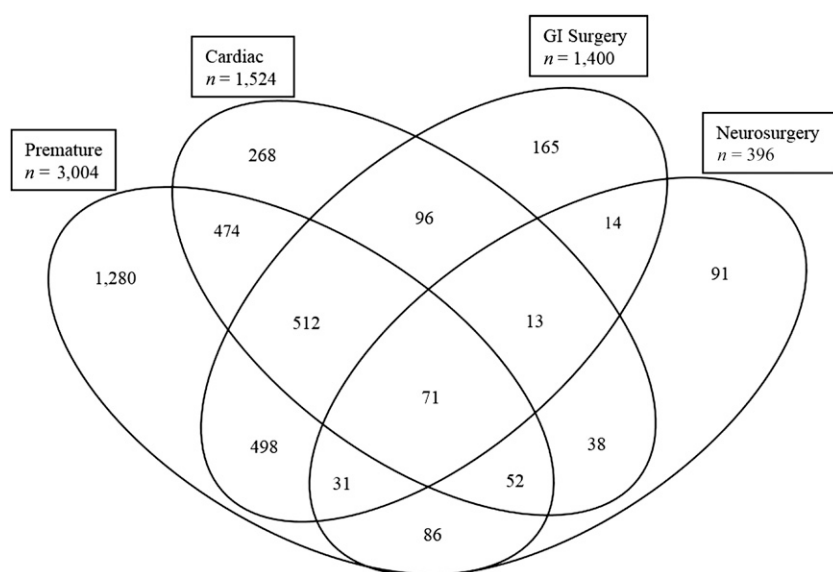


FIGURE 2
Distribution of patients in each subpopulation.

information was included, and its prediction improved as the DTD became smaller (closer to discharge date). Three of the 4 subpopulations and all patients combined performed

very similarly. The only population on which the model consistently underperformed was the NS population, for 2 possible reasons. First, the NS population was the

smallest cohort by far, and therefore the model may not have had enough patients on which to train. Second, the NS population may be very different clinically from the other patients seen in the NICU, and their readiness for discharge may not be captured in the features extracted for this model.

When we broke the most important features down by subpopulation and DTD, the features remained surprisingly consistent across the subpopulations and DTD. This result was unexpected because we thought that different subpopulations 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&Bs, 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 (eg, oral amount divided by the oral amount plus the tube fed amount) was the best-performing or nearly the best-performing feature across populations and DTD values. For 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 >90%. At 10 DTD this feature ranks 20th in importance, but at 2 DTD this feature has advanced to third place. This indicates that consuming most 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 from the hospital in the next 2 to 10 days. However, it may not always be practical or possible to include all these features into a decision support tool to construct this predictive

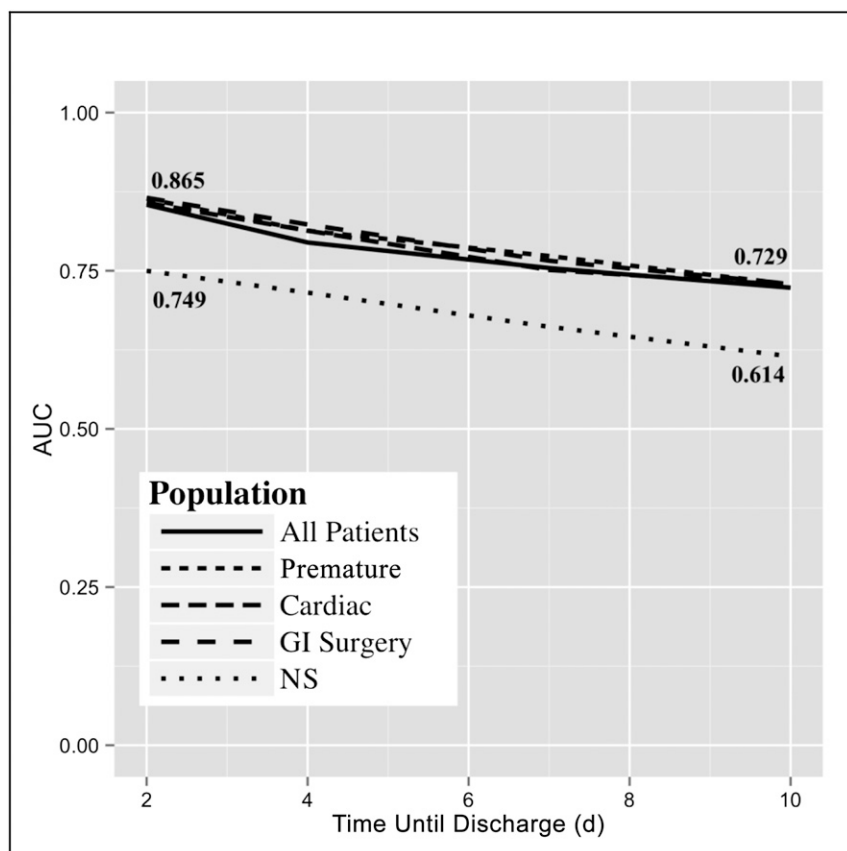


FIGURE 3
AUC for each patient subpopulation for all features.

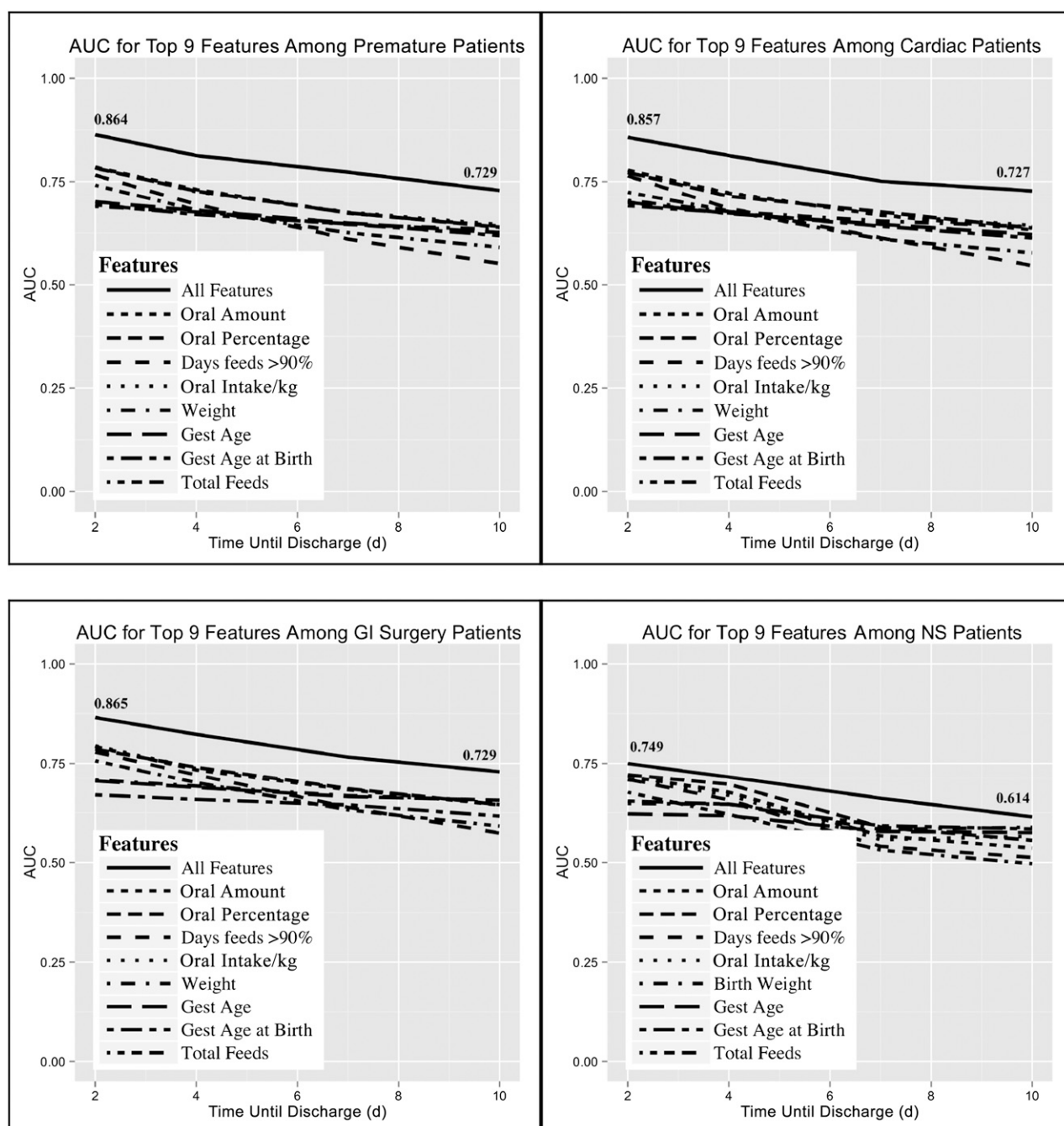


FIGURE 4

The 9 most predictive features for each subpopulation. A single patient may be represented in >1 subpopulation.

model to alert staff of impending discharges. One beneficial aspect 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 determine which patients are nearing discharge. For example, a very simple decision tree

can be constructed from only 2 features (Fig 5). This tree is based on data from all patients, 2 features (oral percentage of feeds and weight), a DTD of 4 days, and a maximum tree depth of 3. The first branch of the tree splits the patients into 2 groups based on whether their oral percentage of feeds is >80%. On the right, the next differentiator is based on weight. If

the patient weighs <1.5 kg, his or her probability of being discharged in the next 4 days is 0.23 (on a scale of 0–1). If the patient weighs between 1.5 and 1.7 kg, his or her probability for discharge in the next 4 days is 0.48. If the patient weighs >1.7 kg and takes >90% of his or her feeds orally, the patient has a 0.81 probability of being discharged in the next 4 days. The

TABLE 2 Top 20 Features in Order of Importance for All Patients for All DTD values

2 DTD		4 DTD		7 DTD		10 DTD	
Feature	AUC	Feature	AUC	Feature	AUC	Feature	AUC
All	0.854	All	0.795	All	0.754	All	0.723
% of oral feeds	0.766	% of oral feeds	0.704	Amount of oral feeds	0.649	% of oral feeds	0.623
Amount of oral feeds	0.764	Amount of oral feeds	0.703	Gestational age birth	0.647	Amount of oral feeds	0.620
No. days oral % >90%	0.753	Amount of oral feeds/wt	0.700	Amount of oral feeds/wt	0.646	Amount of oral feeds/wt	0.620
Amount of oral feeds/wt	0.750	No. days oral % >90%	0.681	% of oral feeds	0.646	Gestational age	0.617
Total feeds	0.720	Gestational age birth	0.678	Birth wt	0.632	Wt	0.617
Gestational age birth	0.707	Gestational age	0.673	Wt	0.632	Gestational age birth	0.609
Birth wt	0.698	Birth wt	0.672	Gestational age	0.631	Birth wt	0.607
Gestational age	0.698	Wt	0.667	No. days off caffeine	0.610	On caffeine	0.594
Wt	0.690	Total feeds	0.652	On caffeine	0.605	No. days off caffeine	0.592
No. days off caffeine	0.643	No. 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
No. days off ventilator	0.624	On caffeine	0.608	No. days oral % >90%	0.589	No. days off oxygen	0.560
No. days off infused medication	0.620	No. days off ventilator	0.605	Cardiac	0.563	On oxygen	0.548
Ratio of wt to birth wt	0.613	No. days off infused medication	0.594	No. days off ventilator	0.562	On ventilator	0.543
On caffeine	0.613	Ratio wt/Birth wt	0.592	Ratio wt/birth wt	0.561	Cardiac	0.542
No. days off oxygen	0.609	Days of life	0.587	No. days off oxygen	0.558	No. days off ventilator	0.537
Days of life	0.604	Cardiac	0.582	No. days off infused medication	0.555	No. of A&Bs	0.535
No. days no A&B	0.601	No. days off oxygen	0.581	Days of life	0.547	No. days oral % >90%	0.534

probabilities for discharge in 4 days for patients at different weights and taking <80% of their feeds orally are listed in the left-side branch.

This simple decision tree has an AUC of 0.843. Although 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.

It is interesting that using all 26 features yields an AUC of 0.865, whereas using only 2 features can yield 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's performance would

be to add more features. The use of trending data (eg, the average amount of feeding increase over a 5-day period) could be beneficial. Another consideration for model improvement would be to predict a range of days until discharge (eg, 3–5 days instead of just 4).

There are several limitations to this study. First, some of the features used in the model are more difficult to obtain than others, and extracting 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 functions built into Vanderbilt's electronic health record. These

functions may not be supported by all electronic health record systems.^{15,16} Third, categorizing hospitalized patients based on ICD-9 codes would be difficult because these codes are not usually available until after discharge. However, as the analysis showed, diagnosis categories added surprisingly little to the prediction model. Should we need our model to differentiate patients, admitting diagnoses could be used. Fourth, our sample could be potentially biased because we did exclude patients if they were missing any progress notes. Although an RF does provide techniques to address missing data, we felt thought excluding these patients was a conservative and appropriate approach.

We trained the model by using actual discharge dates. This limitation worked against us because 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 who had delayed discharges for nonmedical reasons. Additionally, once fully implemented our model might predict discharge too early, which could result in premature expectations of parents and possible wasted effort.

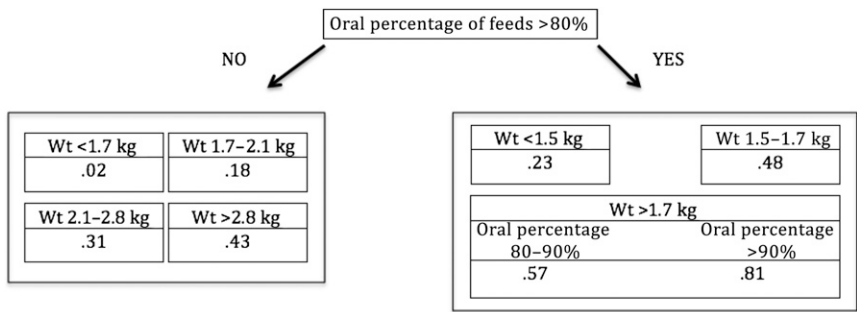


FIGURE 5 A simple decision tree demonstrating how 2 features can be used to create an accurate discharge prediction model. The fraction in each cell denotes the probability of discharge in the next 4 days. This tree has an AUC of 0.843.

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

CONCLUSIONS

A supervised machine learning approach using an RF classifier accurately predicts which patients will be discharged from the NICU in

the next 2 to 10 days. Running our model daily with the most recent progress note data will identify which patients are close to being medically ready for discharge and may alert the clinical staff through indicators in the electronic medical record. This method would allow more timely discharge planning and has the potential to prevent delayed discharges for nonmedical reasons.

ACKNOWLEDGMENT

The authors appreciate the Research Derivative team at Vanderbilt University for their assistance in retrieving data. The publication described was supported by CTSA award No. UL1TR000445 from the National

Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

ABBREVIATIONS

A&B: apnea and bradycardia event
AUC: area under the curve
DTD: days to discharge
GI: gastrointestinal
ICD-9: International Classification of Diseases, Ninth Revision
LOS: length of stay
NS: neurosurgical
RF: random forest

FINANCIAL DISCLOSURE: Dr Lehmann serves in a part-time role at the American Academy of Pediatrics. He has 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 indicated he has no financial relationships relevant to this article to disclose.

FUNDING: National Library of Medicine training grant 5T15LM007450-13.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. Bockli K, Andrews B, Pellerite M, Meadow W. Trends and challenges in United States neonatal intensive care units follow-up clinics. *J Perinatol*. 2014;34(1):71–74
2. Challis D, Hughes J, Xie C, Jolley D. An examination of factors influencing delayed discharge of older people from hospital. *Int J Geriatr Psychiatry*. 2014; 29(2):160–168
3. Victor CR, Healy J, Thomas A, Seargeant J. Older patients and delayed discharge from hospital. *Health Soc Care Community*. 2000;8(6):443–452
4. Szubski CR, Tellez A, Klika AK, 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):e46–e53
5. Marcin JP, Slonim AD, Pollack MM, Ruttimann UE. Long-stay patients in the pediatric intensive care unit. *Crit Care Med*. 2001;29(3):652–657
6. Edwards JD, Houtrow AJ, Vasilevskis EE, 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):2196–2203
7. Ruttimann UE, Pollack MM. Variability in duration of stay in pediatric intensive care units: a multiinstitutional study. *J Pediatr*. 1996;128(1):35–44
8. Powell PJ, Powell CV, Hollis S, Robinson SJ. When will my baby go home? *Arch Dis Child*. 1992;67(10 spec no):1214–1216
9. Bannwart DC, Rebello CM, Sadeck LS, Pontes MD, Ramos JL, Leone CR. Prediction of length of hospital stay in neonatal units for very low birth weight infants. *J Perinatol*. 1999;19(2):92–96
10. Lee SK, McMillan DD, Ohlsson A, et al. Variations in practice and outcomes in the Canadian NICU network: 1996–1997. *Pediatrics*. 2000;106(5):1070–1079
11. Lee HC, Bennett MV, Schulman J, Gould JB. Accounting for variation in length of NICU stay for extremely low birth weight infants. *J Perinatol*. 2013;33(11):872–876
12. Levin SR, Harley ET, Fackler JC, et al. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. *Crit Care Med*. 2012;40(11):3058–3064
13. Sci-Kit Learn. 2014. Available at: <http://scikit-learn.org/stable/index.html>
14. Koppel R, Lehmann CU. Implications of an emerging EHR monoculture for hospitals and healthcare systems. *J Am Med Inform Assoc*. 2015;22(2):465–471
15. Kim GR, Lehmann CU; Council on Clinical Information Technology. Pediatric aspects of inpatient health information technology systems [published correction appears in *Pediatrics*. 2009; 123(2):604]. *Pediatrics*. 2008;122(6). Available at: www.pediatrics.org/cgi/content/full/122/6/e1287
16. Lehmann CU; Council on Clinical Information Technology. Pediatric aspects of inpatient health information technology systems. *Pediatrics*. 2015; 135(3). Available at: www.pediatrics.org/cgi/content/full/135/3/e756

APPENDIX

ICD-9 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

Continued

ICD-9 Code	Description	Category
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 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 g	Premature
765.01	Disorders relating to extreme immaturity of infant, <500 g	Premature
362.24	Retinopathy of prematurity, stage 2	Premature
779.7	Periventricular leukomalacia	Premature
764.95	Unspecified fetal growth retardation, 1250–1499 g	Premature
765	Disorders relating to extreme immaturity of infant, wt unspecified	Premature
764.92	Unspecified fetal growth retardation, 500–749 g	Premature
772.13	Intraventricular hemorrhage of fetus or newborn, grade III	Premature
765.02	Disorders relating to extreme immaturity of infant, 500–749 g	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 wk	Premature
765.17	Disorders relating to other preterm infants, 1750–1999 g	Premature
765.16	Disorders relating to other preterm infants, 1500–1749 g	Premature
765.15	Disorders relating to other preterm infants, 1250–1499 g	Premature

Continued

ICD-9 Code	Description	Category
765.18	Disorders relating to other preterm infants, 2000–2499 g	Premature
765.22	Disorders related to weeks of gestation completed, 24 wk	Premature
765.24	Disorders related to weeks of gestation completed, 27–28 wk	Premature
765.25	Disorders related to weeks of gestation completed, 29–30 wk	Premature
776.6	Anemia of prematurity	Premature
765.27	Disorders related to weeks of gestation completed, 33–34 wk	Premature
765.03	Disorders relating to extreme immaturity of infant, 750–999 g	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 g	Premature
765.13	Disorders relating to other preterm infants, 750–999 g	Premature
765.1	Disorders relating to other preterm infants, wt unspecified	Premature
765.26	Disorders related to weeks of gestation completed, 31–32 wk	Premature

ECMO, extracorporeal membrane oxygenation; PPH, persistent pulmonary hypertension.

Predicting Discharge Dates From the NICU Using Progress Note Data

Michael W. Temple, Christoph U. Lehmann and Daniel Fabbri

Pediatrics 2015;136:e395; originally published online July 27, 2015;

DOI: 10.1542/peds.2015-0456

Updated Information & Services	including high resolution figures, can be found at: /content/136/2/e395.full.html
References	This article cites 15 articles, 5 of which can be accessed free at: /content/136/2/e395.full.html#ref-list-1
Citations	This article has been cited by 1 HighWire-hosted articles: /content/136/2/e395.full.html#related-urls
Post-Publication Peer Reviews (P³Rs)	2 P ³ Rs have been posted to this article /cgi/eletters/136/2/e395
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Fetus/Newborn Infant /cgi/collection/fetus:newborn_infant_sub Health Information Technology /cgi/collection/health_information_technology_sub Biomedical Informatics /cgi/collection/biomedical_informatics_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: /site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: /site/misc/reprints.xhtml

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2015 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Predicting Discharge Dates From the NICU Using Progress Note Data

Michael W. Temple, Christoph U. Lehmann and Daniel Fabbri

Pediatrics 2015;136:e395; originally published online July 27, 2015;

DOI: 10.1542/peds.2015-0456

The online version of this article, along with updated information and services, is located on the World Wide Web at:

</content/136/2/e395.full.html>

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2015 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

