



# Machine learning models to predict and benchmark PICU length of stay with application to children with critical bronchiolitis

Colin M. Rogerson MD, MPH<sup>1</sup>  | Julia A. Heneghan MD<sup>2</sup> |  
 Joseph G. Kohne MD, MSc<sup>3,4</sup> | Denise M. Goodman MD, MS<sup>5</sup> |  
 Katherine N. Slain DO<sup>6</sup> | Cara A. Cecil MD, MS<sup>5</sup>  | Jason M. Kane MD, MS<sup>7</sup> |  
 Matt Hall PhD<sup>8</sup>

<sup>1</sup>Division of Pediatric Critical Care, Indiana University School of Medicine, Indianapolis, Indiana, USA

<sup>2</sup>Division of Pediatric Critical Care, University of Minnesota Masonic Children's Hospital, Minneapolis, Minnesota, USA

<sup>3</sup>Department of Pediatrics, Division of Critical Care Medicine, University of Michigan, Ann Arbor, Michigan, USA

<sup>4</sup>Susan B. Meister Child Health Evaluation and Research Center, University of Michigan School of Medicine, Ann Arbor, Michigan, USA

<sup>5</sup>Northwestern University Feinberg School of Medicine and Ann & Robert H. Lurie Children's Hospital of Chicago, Division of Pediatric Critical Care Medicine, Chicago, Illinois, USA

<sup>6</sup>Division of Pediatric Critical Care, University Hospitals Rainbow Babies & Children's Hospital, Cleveland, Ohio, USA

<sup>7</sup>Department of Pediatrics, Section of Pediatric Critical Care Medicine, Comer Children's Hospital, University of Chicago, Chicago, Illinois, USA

<sup>8</sup>Children's Hospital Association, Lenexa, Kansas, USA

## Correspondence

Colin M. Rogerson, MD, MPH, Riley Hospital for Children at Indiana University Health, 705 Riley Hospital Dr, Phase 2, Room 4900, Indianapolis, IN 46202, USA.  
 Email: [crogerso@iupui.edu](mailto:crogerso@iupui.edu)

## Abstract

**Objective:** To create models for prediction and benchmarking of pediatric intensive care unit (PICU) length of stay (LOS) for patients with critical bronchiolitis.

**Hypothesis:** We hypothesize that machine learning models applied to an administrative database will be able to accurately predict and benchmark the PICU LOS for critical bronchiolitis.

**Design:** Retrospective cohort study.

**Patients:** All patients less than 24-month-old admitted to the PICU with a diagnosis of bronchiolitis in the Pediatric Health Information Systems (PHIS) Database from 2016 to 2019.

**Methodology:** Two random forest models were developed to predict the PICU LOS. Model 1 was developed for benchmarking using all data available in the PHIS database for the hospitalization. Model 2 was developed for prediction using only data available on hospital admission. Models were evaluated using  $R^2$  values, mean standard error (MSE), and the observed to expected ratio (O/E), which is the total observed LOS divided by the total predicted LOS from the model.

**Results:** The models were trained on 13,838 patients admitted from 2016 to 2018 and validated on 5254 patients admitted in 2019. While Model 1 had superior  $R^2$  (0.51 vs. 0.10) and (MSE) (0.21 vs. 0.37) values compared to Model 2, the O/E ratios were similar (1.18 vs. 1.20). Institutional median O/E (LOS) ratio was 1.01 (IQR 0.90–1.09) with wide variability present between institutions.

**Conclusions:** Machine learning models developed using an administrative database were able to predict and benchmark the length of PICU stay for patients with critical bronchiolitis.

## KEYWORDS

bronchiolitis, informatics, machine learning, pediatric critical care, pediatrics

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Pediatric Pulmonology* published by Wiley Periodicals LLC.

## 1 | INTRODUCTION

Hospital length of stay (LOS) is a useful administrative metric to assess a hospital's efficiency and provides insights into hospital workflow and throughput. As a quality measure, hospital LOS is reported to Centers for Medicare and Medicaid Services as part of the Hospital Inpatient Quality Reporting Program, and to the Agency for Healthcare Research and Quality.<sup>1,2</sup> Prior studies have attempted to use large clinical and administrative databases to accurately predict hospital LOS using both traditional statistical methods,<sup>3,4</sup> and machine learning methods.<sup>5–12</sup> While previous machine learning models have been applied to various specialties and clinical settings, few have been built specifically to predict LOS in the pediatric intensive care unit (PICU).

The ability to assess a hospital's performance based on actual patient LOS compared to predicted LOS is valuable to hospital administrators, health services researchers, and quality improvement specialists. The ratio of observed-to-expected LOS allows for benchmarking across different hospitals and can enable direct comparisons.<sup>13</sup> To accurately model a standardized LOS ratio that can be used for comparisons, a common disease condition seen frequently across multiple institutions should be used. In the United States, viral bronchiolitis is the most common clinical condition requiring pediatric hospitalization, costing \$734 million in 2016 and with costs increasing over time.<sup>14–16</sup> Children with critical bronchiolitis also frequently require admission to the PICU. For patients that require PICU care, the PICU LOS accounts for 15%–60% of total hospital days in academic children's hospitals in the United States.<sup>17</sup> For these reasons, we chose to use critical bronchiolitis as the population of interest to develop both a predictive model and benchmarking model for PICU length of stay. Pediatric hospital LOS prediction by machine learning methodology has been advocated for in the published literature<sup>6</sup> as it allows for the selection of variables influential in the prediction of patient LOS without a priori identification by clinicians.

The primary objective of this study was to create accurate models to both predict and benchmark the PICU LOS for pediatric patients admitted with critical bronchiolitis using machine learning methods. The secondary objective was to evaluate the effects of using data at different points in time during the hospital stay on model performance. We hypothesized that machine learning models developed using administrative data obtained on hospital admission could accurately predict and benchmark PICU LOS in this population.

## 2 | MATERIALS AND METHODS

This was a retrospective cohort study using data from a large, administrative database. This study was reviewed by the Indiana University Institutional Review Board and granted exemption (IRB #13768). Data for this study was obtained from the Pediatric Health Information System (PHIS) database. PHIS is an administrative database sponsored by the Children's Hospital Association (Lenexa,

KS) and contains patient-level information from inpatient, ambulatory surgery, emergency department, and observation unit encounters. In addition to demographics, diagnoses, and procedures commonly collected in administrative databases, PHIS also contains daily billing information for each patient. The PHIS database contains healthcare data from 49 pediatric hospitals in the United States. The PHIS database applies several validated levels of data quality and reliability checks in the collection process. All patient data is de-identified at the time of data submission.

The study population included all admissions of pediatric patients with a primary diagnosis of bronchiolitis admitted to the PICU between 2016 and 2019. The diagnosis of bronchiolitis was defined as age less than 24 months and an All Patients Refined Diagnosis Related Groups (APR DRG) code 138, which specifies bronchiolitis and respiratory syncytial virus (RSV) pneumonia. PICU admission was defined using the presence of a PICU room charge on a given day. Patients who were initially admitted to the hospital floor and subsequently transferred to the PICU were also included.

Random forest (RF) models are a specific type of classification and regression tree (CART) models. These supervised machine learning models use bootstrapped, aggregated sampling techniques to reduce variability in the final model and increase predictive accuracy.<sup>18</sup> At a high level, RFs combine many individual CART models that are generated from resampled data into a single model. Two separate RF models were constructed in this study. Model 1 was developed for benchmarking by using all available data in the database for the hospital encounter from day of admission through the last PICU day. This includes patient demographics (age, sex, insurance payer, admission source), the presence of complex chronic conditions<sup>19</sup> (CCCs), the number of hospitalizations and total number of hospital days in the prior year, diagnoses present on admission grouped using the Agency for Healthcare Research and Quality's Clinical Classification System algorithm (CCS),<sup>20</sup> procedures performed, medications ordered, laboratory tests ordered, and radiology tests ordered. Model 2 was developed for prediction using only those variables present at time of hospital admission (either floor or PICU admission), and included patient demographics, CCCs present, the number of hospitalizations and total number of hospital days in the prior year, and diagnoses present on admission. No procedural or medication data was used in this model. All included variables are required fields in PHIS except for sex and payor. Although PHIS contains data related to LOS, no variables related to LOS were included in the model development.

Both models used data from years 2016–2018 as a training data set, and data from 2019 as a validation data set. Models were constructed using 500 bootstrap aggregate samples. The outcome of interest was PICU LOS in days. LOS was log-transformed to reduce the influence of outliers. Regression trees use multiple decision nodes to predict the outcome of interest based on the input variables. Nodal decisions that correctly predict the outcome of interest increase the purity of the model. Thus, the variables that provide the greatest increase in the ability of the model to accurately predict the outcome of interest increase the node purity and are selected as the variables

of highest importance. The variable importance is reported using the mean increase in node purity. Model accuracy was assessed using  $R^2$  (a measure of the proportion of variance in the outcome that is explained by the model), mean squared error (MSE) (a direct measurement of how far the observed outcome is from that predicted by the model), and observed to expected ratio (the observed outcome divided by the model-predicted outcome). Since predicted values for the LOS from the RF were on the log-scale, they were exponentiated and rounded up to the next whole integer to match the scale of the data available in PHIS. We also performed a sensitivity analysis by testing both models on patients with shorter LOS (1–2 days), and longer LOS (>2 days). Categorical comparisons were made using  $\chi^2$  tests; continuous comparisons were made using Wilcoxon Rank-Sum tests. Hypothesis tests were considered statistically significant at an  $\alpha$  level of 0.05. Statistical analyses were performed using SAS v.9.4 (SAS Institute) and R 3.2.0 using the “randomForest” package.

## 3 | RESULTS

### 3.1 | Patient inclusion

Data for this study was obtained from the 47 PHIS-participating hospitals which provided data throughout the study period and included 19,092 admissions across the validation and training cohorts. The training cohort (2016–2018) included 13,838 admissions, and the validation cohort (2019) included 5254 admissions. Sex was missing for 12 (0.06%) patients in the training cohort who were excluded. Payor was missing in 191 discharges between both cohorts and was classified as “Other.” Table 1 illustrates the comparison between the validation and training cohorts. There were no significant differences in patient age, sex, payor, or total PICU LOS. There were significant differences in race, hospital LOS, days of PICU LOS, and mortality between the cohorts. The training cohort had longer hospital LOS, differences in days of PICU stay, higher mortality, a higher percentage of government payors, a higher proportion of non-Hispanic white patients and a lower proportion of Hispanic patients compared to the validation cohort.

### 3.2 | Model 1

Model 1 was created for benchmarking with all available data for the hospital encounter from hospital admission through the last PICU day. This model included a total of 2084 variables (189 CCS diagnosis groups, 80 CCS procedural groups, 12 CCC groups, admission source, age in days, number of prior encounters in the past year, payor, sex, day of PICU admission, 1796 billed resources). This model achieved an  $R^2$  value of 0.51, an MSE of 0.21, and an observed to expected ratio of 1.18 (95% CI: 1.17, 1.20). This means that for every 5 days the model expected a patient to stay in the PICU, the patient, on

average, stayed 6 days. The top five variables with the highest influence in this model were: furosemide, supine abdominal X-ray, acetaminophen, age, and chest physical therapy (Table 2). In the sensitivity analysis, this model performed better on patients with longer LOS (>2 days) ( $R^2 = 0.40$ , MSE = 0.10, O/E = 1.09) compared to shorter LOS (1–2 days) ( $R^2 = 0.18$ , MSE = 0.10, O/E = 1.21). The model calibration plot is included as Supporting Information: Figure 1.

### 3.3 | Model 2

Model 2 was created for prediction using only administrative data available at the time of hospital admission. This model included a total of 207 variables (189 CCS dx groups, 12 CCC groups, admission source, age in days, number of prior encounters, payor, sex). This model achieved an  $R^2$  value of 0.10, an MSE of 0.37, and an observed to expected ratio of 1.20 (95% CI: 1.18, 1.23). The top five variables that were the most influential on the model's prediction were: age, the number of previous hospital days in the last year, an admission diagnosis of pneumonia, hospital admission source, and payor (Table 2). Comparing these results to those of Model 1, while the  $R^2$  and MSE values demonstrate a substantial decrease in model performance, the observed to expected ratio was essentially unchanged. In the sensitivity analysis, this model also performed better on patients with longer LOS (>2 days) ( $R^2 = 0.08$ , MSE = 0.15, O/E = 1.07) compared to shorter LOS (1–2 days) ( $R^2 = 0.05$ , MSE = 0.13, O/E = 1.28). The model calibration plot is included as Supporting Information: Figure 2.

### 3.4 | Institutional variability

The observed to expected ratios for Model 1 were obtained for each PHIS-contributing institution represented in the 2019 cohort. These ratios are displayed in Figure 1 for visual comparison. The median for all institutions was 1.01 with an interquartile range of 0.90 to 1.09. The largest observed-to-expected ratio was 1.80 (meaning an expected stay of 5 days would have actually been 9 days), and the smallest was 0.63 (actual stay 3 days).

## 4 | DISCUSSION

This retrospective cohort study using an administrative pediatric database demonstrates the use of machine learning methods to both benchmark and predict the PICU LOS for patients admitted to the PICU with critical bronchiolitis. Model 1 used all available inpatient data from day of hospital admission through the last PICU day. Model 2 used only data available at hospital admission, and while it had lower model performance metrics, its observed to expected length of stay ratio was comparable to that of Model 1. An institutional comparison revealed substantial differences in the PICU length of stay for critical bronchiolitis by institution (Figure 1).

**TABLE 1** Demographics and clinical characteristics of the training and validation cohorts.

		Overall	Training (2016–2018)	Validation (2019)	p Value
N		19092	13838	5254	
Age (months)	0–1	3415 (17.9)	2496 (18.0)	919 (17.5)	0.808
	2–5	5690 (29.8)	4128 (29.8)	1562 (29.7)	
	6–11	4845 (25.4)	3503 (25.3)	1342 (25.5)	
	12–24	5142 (26.9)	3711 (26.8)	1431 (27.2)	
Sex	Male	11342 (59.4)	8228 (59.5)	3114 (59.3)	0.776
	Female	7738 (40.6)	5599 (40.5)	2139 (40.7)	
Race	Non-Hisp White	8431 (44.2)	2387 (45.4)	6044 (43.7)	0.031
	Non-Hisp Black	4447 (23.3)	1219 (23.2)	3228 (23.3)	
	Hispanic	3501 (18.3)	910 (17.3)	2591 (18.7)	
	Asian	562 (2.9)	171 (3.3)	391 (2.8)	
	Other	2151 (11.3)	567 (10.8)	1584 (11.4)	
Payor	Government	12155 (63.7)	8882 (64.2)	3273 (62.3)	0.053
	Private	6092 (31.9)	4351 (31.4)	1741 (33.1)	
	Other	845 (4.4)	605 (4.4)	240 (4.6)	
Mortality		14 (0.1)	14 (0.1)	0 (0.0)	0.021
Total PICU LOS-days	Geo mean (std dev)	2.2 (1.9)	2.2 (1.9)	2.1 (1.9)	0.058
Total LOS-days	Geo mean (std dev)	4.0 (1.9)	4.1 (1.9)	3.8 (1.9)	<0.001
Days of PICU stay	1	881 (4.6)	260 (4.9)	621 (4.5)	<0.001
	2	3317 (17.4)	987 (18.8)	2330 (16.8)	
	3	3998 (20.9)	1122 (21.4)	2876 (20.8)	
	4	3148 (16.5)	886 (16.9)	2262 (16.3)	
	5	2215 (11.6)	587 (11.2)	1628 (11.8)	
	6	1554 (8.1)	418 (8)	1136 (8.2)	
	7+	3979 (20.8)	994 (18.9)	2985 (21.6)	

Abbreviations: LOS, length of stay; PICU, pediatric intensive care unit.

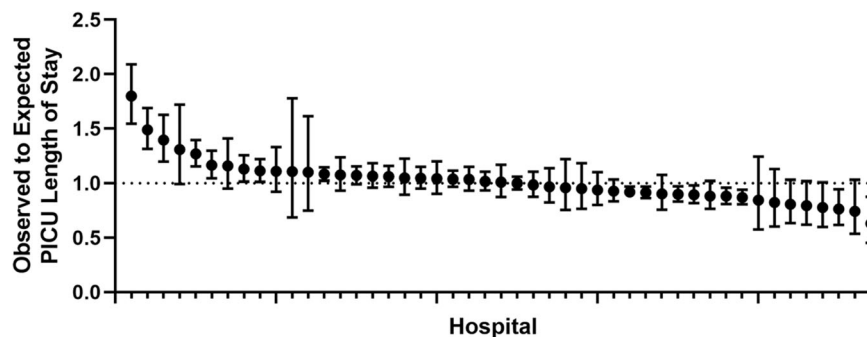
The ability to predict patient LOS is a useful tool for hospital administrators, quality improvement scientists, and clinicians. It allows administrators to facilitate hospital and unit-specific workflow and staffing needs. It allows quality improvement scientists to monitor institutional efficiency and compare it to similar institutions. Accurate LOS prediction also allows clinicians to provide anticipatory guidance to patients and families and align expectations as to duration of care. While classification algorithms can potentially provide more interpretable accuracy measurements and results (such as prediction of prolonged ICU stay with metrics such as sensitivity, specificity, positive and negative predictive values), regression algorithms such as those presented in this study can provide more specific predictions. The models developed in this study can predict the number of days a patient with critical viral bronchiolitis will stay in the PICU with an average error of 1 day for every 5. While mathematical measures of regression model accuracy ( $R^2$  and MSE)

are important for model comparisons, they do not translate well into actual practice or clinical usefulness. The observed to expected ratio provides a more usable measure of regression model accuracy which can be understood and compared as a hospital performance metric.

Model 1 used all available data in the PHIS administrative database from hospital admission through the last PICU day for each included patient. This model produced superior  $R^2$  and MSE values, but only improved the observed to expected LOS ratio by 0.02. While this model overall is superior to Model 2 in terms of accuracy metrics, the need for substantial additional data may limit its applicability for providers at the bedside and would not provide a prospective LOS prediction. Quality improvement specialists may find a use for this type of model, however. Administrative databases such as PHIS can provide institutional comparisons of these metrics, allowing institutional representatives to retrospectively compare and benchmark their data to peer institutions. High-performing institutions can be

**TABLE 2** Top 10 variable importance measurements for each model.

<b>Model 1: Total Hospital Encounter Data Model Validation: <math>R^2 = 0.51</math>, Mean Square Error = 0.21, Mean(Obs/Exp) = 1.18</b>		
Variable	Mean increase in node purity	Percent increase in mean squared error
Abdomen supine X-ray	525.8	118.7
Furosemide	434.3	67.6
Age in days	268.6	43.9
Acetaminophen (APAP) (N-acetyl-P-aminophenol)	259.1	87.2
Mechanical ventilation	207.5	56.0
Chest physical therapy	167.6	49.4
Midazolam HCl	78.8	13.8
Other specified dietary service	62.9	36.5
Respiratory intubation and mechanical ventilation (PR CCS 216)	60.6	20.5
Number of hospital days within the prior 1 year	59.9	11.8
<b>Model 2: Hospital Admission Data Model Validation: <math>R^2 = 0.10</math>, Mean Square Error = 0.37, Mean(Obs/Exp) = 1.20</b>		
Age in days	1239.6	72.5
Number of hospital days within the prior 1 year	297.5	30.7
Pneumonia (CCS 122)	176.9	59.9
Admission source	175.4	4.6
Payor	171.4	3.4
Number of hospitalizations within the prior 1 year	132.7	18.4
Sex	127.1	6.7
Viral infection (CCS 7)	111.4	8.0
Fluid and electrolyte disorders (CCS 55)	96.0	8.4
Cardiovascular CCC	85.0	17.5

**FIGURE 1** Observed to expected pediatric intensive care unit length of stay ratio by hospital. The y-axis illustrates the observed to expected ratio for the 2019 validation cohort, with the center line at 1.0. The x-axis represents individual hospitals ordered from the highest calculated observed to expected ratio to the lowest.

identified, and specific processes can be defined and implemented to reduce variability and improve the healthcare quality at all participating centers. This is the approach taken when using APR-DRG codes to create “expected” LOS values for specific diagnostic codes.<sup>21</sup> While the PHIS measured LOS is rounded to whole days and may be less useful in measuring small improvements in the management of critical bronchiolitis, the approach to machine learning modeling in this study could be used as a proof of concept and applied to more heterogeneous processes with more variable hospital courses. Model 2 used only data available at the time of hospital admission to predict the PICU LOS. While the  $R^2$  and MSE values are generally low and substantially lower than those provided by Model 1, the observed to expected PICU LOS ratio is essentially unchanged. This type of model would be more useful in the clinical setting, where a provider will not know the expansive list of orders or procedures a patient will experience while they are caring for the patient, but could still give prognostic counsel to the patient or patient family. It would also be useful to administrators attempting to manage hospital and unit workflows for staffing needs and bed-planning purposes, as they would potentially have better predictive data early in a patient's hospital course. The lower accuracy metrics of this model limit the ability to predict person-to-person variation in PICU LOS but could provide useful aggregate data for hospital workflow purposes. In the sensitivity analysis, both models performed better on patients with longer LOS than with shorter LOS. One explanation could be that variability in short stay patients is harder to predict with the available variables and could be due to hospital system differences such as PICU step-down units. Both models would be more clinically and administratively useful in predicting longer LOS accurately.

RF models rank the input variables in order of importance in the final prediction. Importantly, while RF models do provide a measurement of variable importance, unlike linear or logistic regression models, they do not provide any measure of directionality for the variables. Model 2 identified several factors that make clinical sense in predicting the acuity of critical bronchiolitis: patient age, number of CCCs present, prior days in the hospital during the past year, and the diagnosis of pneumonia. However, the importance of patient origin and insurance payor are surprising. It is possible that these are surrogate markers for delayed presentation to the hospital, or intubation in an adult emergency department that has less experience

with critical bronchiolitis. The factors identified in Model 1 are even more surprising. A few make intuitive sense: patient age and mechanical ventilation. However, others included: furosemide, abdominal X-ray, chest physiotherapy, and acetaminophen. It is also likely in this case that these are surrogate markers for patient acuity. For example, patients with acute kidney injury, fluid overload, or a cardiac comorbidity could have need for a diuretic. An order for an abdominal radiograph may be needed in children that are unable to eat by mouth due to respiratory distress and undergo imaging to verify nasogastric tube placement. The need for acetaminophen could be a marker for persistent fevers and a potential bacterial superinfection. The primary goal of random forest models, however, is not to provide variable inference, but to provide accurate prediction, or in the case of Model 1, benchmarking.

This study has several limitations. First, PHIS is an administrative database consisting primarily of large, academic children's hospitals, and therefore may not represent the general population of children needing critical care for bronchiolitis. The data obtained from PHIS is limited to administrative discharge data without clinical information beyond what can be captured from billing and claims data derived from International Classification of Diseases and procedure codes. Because of this, there may be clinical differences in the patient populations not captured by the administrative data set. PICU LOS is also measured in whole days in PHIS and may limit the ability to detect small improvements on the order of hours which may be important in critical bronchiolitis. Also, specific institutions may vary in what therapeutics must be provided in an intensive care unit versus what can be used on the floor. These variations will likely affect the PICU LOS for this patient population, and this data is not available in PHIS. As tertiary care academic hospitals, these institutions likely care for more complex patients than the general community PICU population, which may lead to longer observed lengths of stay.<sup>22</sup> Both models incorporated patient sex, which may lead to difficulty with future implementation. Also, the training and validation cohorts had differences between them. While statistically significant, these differences were small and may not be clinically relevant. These data were also used only for viral bronchiolitis in children under 2 years of age and will have limited generalizability to length of stay predictions for other patient populations or clinical conditions that may be more heterogeneous. This study is retrospective, and thus any reported associations cannot be used to demonstrate causal pathways.

## 5 | CONCLUSION

Machine learning models developed from administrative data were able to predict and benchmark the length of stay in the PICU for patients with critical bronchiolitis. These models have potential use for hospital administrators, quality improvement scientists, and bedside clinicians, to help with benchmarking and monitoring quality of care, particularly if these models can be broadened to include more heterogeneous groups of clinical disorders.

## AUTHOR CONTRIBUTIONS

**Colin M. Rogerson:** Conceptualization; investigation; writing—original draft; methodology; validation; writing—review & editing; project administration; supervision; visualization. **Julia A. Heneghan:** Conceptualization; investigation; methodology; validation; visualization; writing—review & editing. **Joseph G. Kohne:** Conceptualization; investigation; validation; methodology; visualization; writing—review & editing. **Denise M. Goodman:** Conceptualization; investigation; methodology; validation; visualization; writing—review & editing; supervision. **Katherine N. Slain:** Conceptualization; investigation; methodology; validation; visualization; writing—review & editing. **Cara A. Cecil:** Conceptualization; investigation; methodology; validation; visualization; writing—review & editing. **Jason M. Kane:** Conceptualization; investigation; methodology; validation; visualization; writing—review & editing. **Matt Hall:** Conceptualization; investigation; writing—original draft; methodology; validation; visualization; writing—review & editing; software; formal analysis; project administration; data curation; supervision; resources.

## ACKNOWLEDGMENTS

Children's Hospital Association. Pediatric Health Information Systems Pediatric Intensive Care Research Node.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## ORCID

Colin M. Rogerson  <http://orcid.org/0000-0001-5251-2399>

Cara A. Cecil  <http://orcid.org/0000-0001-8083-2365>

## REFERENCES

1. Bueno H, Ross JS, Wang Y, et al. Trends in length of stay and short-term outcomes among Medicare patients hospitalized for heart failure, 1993–2006. *JAMA*. 2010;303(21):2141–2147. doi:10.1001/jama.2010.748
2. Ma X, Si Y, Wang Z, Wang Y. Length of stay prediction for ICU patients using individualized single classification algorithm. *Comput Methods Programs Biomed*. 2020;186:105224. doi:10.1016/j.cmpb.2019.105224
3. Wilk M, Marsh DWR, De Freitas S, Prowle J. Predicting length of stay in hospital using electronic records available at the time of admission. *Stud Health Technol Inform*. 2020;270:377–381. doi:10.3233/shti200186
4. Doctoroff L, Herzig SJ. Predicting patients at risk for prolonged hospital stays. *Med Care*. 2020;58(9):778–784. doi:10.1097/mlr.0000000000001345
5. Lequartier V, Wang T, Fondrevelle J, Augusto V, Duclos A. Hospital length of stay prediction methods: a systematic review. *Med Care*. 2021;59(10):929–938. doi:10.1097/mlr.0000000000001596
6. Medeiros NB, Fogliatto FS, Rocha MK, Tortorella GL. Forecasting the length-of-stay of pediatric patients in hospitals: a scoping

- review. *BMC Health Serv Res.* 2021;21(1):938. doi:10.1186/s12913-021-06912-4
7. Bacchi S, Tan Y, Oakden-Rayner L, Jannes J, Kleinig T, Koblar S. Machine learning in the prediction of medical inpatient length of stay. *Intern Med J.* 2022;52(2):176-185. doi:10.1111/imj.14962
  8. Wu J, Lin Y, Li P, Hu Y, Zhang L, Kong G. Predicting prolonged length of ICU stay through machine learning. *Diagnostics.* 2021; 11(12):2242. doi:10.3390/diagnostics11122242
  9. Lin WT, Wu TY, Chen YJ, Chang YS, Lin CH, Lin YJ. Predicting in-hospital length of stay for very-low-birth-weight preterm infants using machine learning techniques. *J Formos Med Assoc.* 2022;121: 1141-1148. doi:10.1016/j.jfma.2021.09.018
  10. Metsker O, Kopanitsa G, Malushko A, et al. Gynecological surgery and machine learning: complications and length of stay prediction. *Stud Health Technol Inform.* 2021;281:575-579. doi:10.3233/shti210236
  11. Alsinglawi B, Alnajjar F, Mubin O, et al. Predicting length of stay for cardiovascular hospitalizations in the intensive care unit: machine learning approach. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020;2020: 5442-5445. doi:10.1109/embc44109.2020.9175889
  12. Pollack MM, Holubkov R, Reeder R, et al. PICU length of stay: factors associated with bed utilization and development of a benchmarking model. *Pediatric Critical Care Medicine.* 2018;19(3): 196-203. doi:10.1097/PCC.0000000000001425
  13. Fujiogi M, Goto T, Yasunaga H, et al. Trends in bronchiolitis hospitalizations in the United States: 2000–2016. *Pediatrics.* 2019;144(6):e20192614. doi:10.1542/peds.2019-2614
  14. Pelletier JH, Au AK, Fuhrman D, Clark RSB, Horvat C. Trends in bronchiolitis ICU admissions and ventilation practices: 2010–2019. *Pediatrics.* 2021;147(6):e2020039115. doi:10.1542/peds.2020-039115
  15. Slain KN, Malay S, Shein SL. Hospital charges associated with critical bronchiolitis from 2009 to 2019. *Pediatr Crit Care Med.* 2022;23(3): 171-180. doi:10.1097/PCC.0000000000002878
  16. Heneghan JA, Rogerson C, Goodman DM, Hall M, Kohne JG, Kane JM. Epidemiology of pediatric critical care admissions in 43 United States children's hospitals, 2014–2019. *Pediatr Crit Care Med.* 2022;23:484-492. doi:10.1097/pcc.0000000000002956
  17. Rigatti SJ. Random forest. *J Insur Med.* 2017;47(1):31-39. doi:10.17849/inm-47-01-31-39.1
  18. Feudtner C, Feinstein JA, Zhong W, Hall M, Dai D. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr.* 2014;14:199. doi:10.1186/1471-2431-14-199
  19. Clinical Classifications Software (CCS) for ICD-10-PCS (beta version). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality. Accessed May 24, 2022. <https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp>
  20. Richardson T, Rodean J, Harris M, Berry J, Gay JC, Hall M. Development of hospitalization resource intensity scores for kids (H-RISK) and comparison across pediatric populations. *J Hosp Med.* 2018;13(9):602-608. doi:10.12788/jhm.2948
  21. Simon TD, Berry J, Feudtner C, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. *Pediatrics.* 2010;126(4):647-655. doi:10.1542/peds.2009-3266
  22. Alghatani K, Ammar N, Rezgui A, Shaban-Nejad A. Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation. *JMIR Med Inform.* 2021;9(5):e21347. doi:10.2196/21347
  23. Slain KN, Shein SL, Stormorken AG, Broberg MCG, Rotta AT. Outcomes of children with critical bronchiolitis living in poor communities. *Clin Pediatr.* 2018;57(9):1027-1032. doi:10.1177/000922817740666
  24. Inagaki K, Blackshear C, Burns PA, Hobbs CV. Racial/ethnic disparities in the incidences of bronchiolitis requiring hospitalization. *Clin Infect Dis.* 2021;72(4):668-674. doi:10.1093/cid/ciaa113

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Rogerson CM, Heneghan JA, Kohne JG, et al. Machine learning models to predict and benchmark PICU length of stay with application to children with critical bronchiolitis. *Pediatric Pulmonology.* 2023;58:1777-1783. doi:10.1002/ppul.26401