Resuscitation 85 (2014) 1065–1071

View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Resuscitation

journal homepage: www.elsevier.com/locate/resuscitation

Developing and evaluating a machine learning based algorithm to predict the need of pediatric intensive care unit transfer for newly hospitalized children

Haijun Zhai [a], Patrick Brady [b,d], Qi Li [a], Todd Lingren [a], Yizhao Ni [a], Derek S. Wheeler [c], Imre Solti [a,d,*]

[a] Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[b] Division of Hospital Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[c] Division of Critical Care Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[d] James M. Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

ARTICLE INFO

Article history:
Received 27 September 2013
Received in revised form 20 February 2014
Accepted 8 April 2014

Keywords:
Clinical status deterioration
Clinical care
PICU
PEWS
Machine learning
EHR

ABSTRACT

Background: Early warning scores (EWS) are designed to identify early clinical deterioration by combining physiologic and/or laboratory measures to generate a quantified score. Current EWS leverage only a small fraction of Electronic Health Record (EHR) content. The planned widespread implementation of EHRs brings the promise of abundant data resources for prediction purposes. The three specific aims of our research are: (1) to develop an EHR-based automated algorithm to predict the need for Pediatric Intensive Care Unit (PICU) transfer in the first 24 h of admission; (2) to evaluate the performance of the new algorithm on a held-out test data set; and (3) to compare the effectiveness of the new algorithm's with those of two published Pediatric Early Warning Scores (PEWS).
Methods: The cases were comprised of 526 encounters with 24-h Pediatric Intensive Care Unit (PICU) transfer. In addition to the cases, we randomly selected 6772 control encounters from 62516 inpatient admissions that were never transferred to the PICU. We used 29 variables in a logistic regression and compared our algorithm against two published PEWS on a held-out test data set.
Results: The logistic regression algorithm achieved 0.849 (95% CI 0.753–0.945) sensitivity, 0.859 (95% CI 0.850–0.868) specificity and 0.912 (95% CI 0.905–0.919) area under the curve (AUC) in the test set. Our algorithm's AUC was significantly higher, by 11.8 and 22.6% in the test set, than two published PEWS.
Conclusion: The novel algorithm achieved higher sensitivity, specificity, and AUC than the two PEWS reported in the literature.
© 2014 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

1. Introduction

Failure to rescue hospitalized patients from complications of disease or treatment is the source of substantial morbidity and death.[1,2] A cardiopulmonary arrest or code outside the intensive care unit (ICU) is a profound consequence of failure to rescue that is associated with a poor prognosis in hospitalized children and adults.[3] As clinical antecedents are present before most codes, rapid response systems (RRS) have been designed, tested, and implemented to detect deterioration early and to rapidly intervene.[4,5] One challenge with RRS is failure to activate or trigger the afferent limb.[6] Early warning scores (EWS) are designed to address this challenge by combining physiologic and/or laboratory measures into a quantified score that can then be linked to clear, expected action such as increased nursing assessments or activation of RRS.[7–18] The most commonly used Pediatric EWS (PEWS) combine scores in 3–7 sub-scales to generate a score between 0 and 26.[12,15,16] Initial development and validation of these scores, which are designed to be tabulated by hand by nurses, occurred before widespread implementation of electronic health records (EHR) and therefore leverage only a small fraction of the EHR content.

The predictive validity of two commonly used PEWS scores[12,15,16] has been examined using the outcome of subsequent transfer to the PICU. The Bedside PEWS is the most extensively validated to date and includes seven components: heart rate, systolic blood pressure, capillary refill time, respiratory

* Corresponding author at: Cincinnati Children's Hospital Medical Center, Division of Biomedical Informatics, 3333 Burnet Avenue, MLC 7024, Cincinnati, OH 45229-3039, USA.
E-mail address: imre.solti@cchmc.org (I. Solti).

http://dx.doi.org/10.1016/j.resuscitation.2014.04.009
0300-9572/© 2014 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

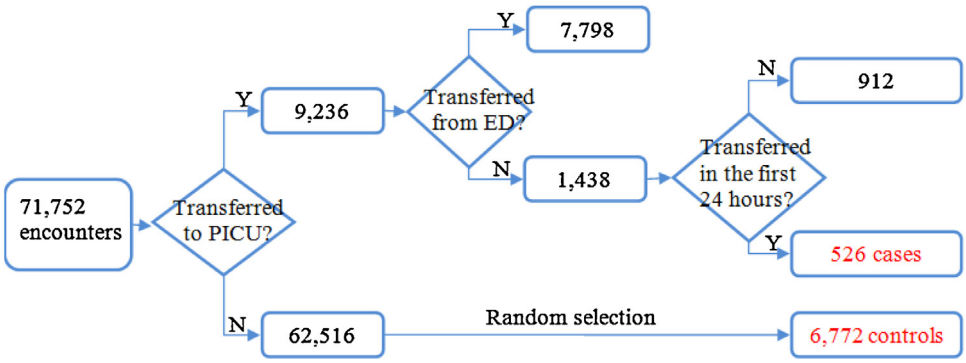

Fig. 1. Steps to generate cases and controls.

rate, respiratory effort, transcutaneous oxygen saturation, and oxygen therapy.[15] A score of 0, 1, 2, or 4 is generated from each category and aggregated to a total score, which has an area under the receiving operating characteristics curve (AUC) of 0.91 in its derivation cohort and AUC of 0.87 and 0.73 in two separate validation cohorts.[12,15,17]

The Monaghan's PEWS used in our institution combines subscores in behavior, cardiovascular, and respiratory domains, with added points for nebulizers ¼ hourly or vomiting following surgery to create a 0–9 overall score. While less extensively validated, this score had AUC of 0.89 when prospectively evaluated.[16] Since an EWS will only succeed in preventing deterioration when it is tied to clear action, each score has cut points where associated algorithms call for specific actions to be taken. The Bedside PEWS has most commonly been studied using a cut point of 8, while the Monaghan's PEWS commonly uses a score >2 for increased nurse and physician evaluation.[15,16]

The planned widespread implementation of EHRs brings the promise of abundant data resources for research purposes via secondary use of EHR data, including better prediction of clinical deterioration.[19] As noted, EHRs and EHR-based research can transform health care delivery through advanced clinical decision support.[20] However, many of the grand challenges in developing clinical decision support are still barely addressed.[21] One of these challenges is to mine large clinical data sets to develop new clinical decision support systems to improve clinical outcomes. In our study we aim to contribute to achieving this exact goal by using the data collected in the EHR during routine clinical care to derive and evaluate a prediction algorithm for PICU transfer for children in acute care wards within the first 24 h of admission.

2. Methods

2.1. Definition of cases and controls

Cincinnati Children's Hospital Medical Center's (CCHMC) Institutional Review Board approved the protocol for our retrospective study. We extracted EHR data that were generated by clinical providers between January 1, 2010 and August 31, 2012. During this period, CCHMC had 71,752 admissions to its inpatient wards. Of these, 1438 admissions were later transferred from the general wards to the PICU. Our unit of analysis was the encounter and not the patient. For each inpatient encounter, we defined the first 24 h of admission as the study period for three reasons. First, we attempted to determine which patients might need more attention and resources at the start of their inpatient stay. Second, as presented below, the PICU transfers that occurred in this scope covered a large percentage of total PICU transfers (i.e., 36.6%). Third, the algorithm developed in this scope could be generalized and tested in other scopes. We identified 526 case and 6772 control encounters (Fig. 1).

Cases and controls were split into two experimental datasets, a training set with 90% of cases (including 473 cases and 473 controls) and a test set with 10% of cases (consisting of 53 cases and 6299 controls). The 119:1 ratio of "no-PICU transfer": "24-h PICU transfer" was maintained in the test set to preserve the generalizability of the study's findings.

2.2. Identification and selection of predictive clinical elements for the machine learning algorithm

We collected over 300,000,000 data points from all 71,752 encounters that occurred between January 1, 2010 and August 31, 2012. The data set included 7587 unique clinical elements as candidate predictors. Through a six-step process (Fig. 2), we selected the predictive clinical elements from this data set.

In the first step, we sorted the clinical elements by their frequency. In the next step we filtered out the elements that were measured in less than 20% of clinical encounters and retained the top 400 most frequent elements. In the third step, a pediatric hospitalist manually reviewed the 400 clinical elements and generated a list of 16 candidate clinical elements with predictive potential. To create independent variables, we collected all measurements for the 16 clinical elements recorded in the EHR until 1 h before the transfer event for cases and measurements recorded in the first 24 h for

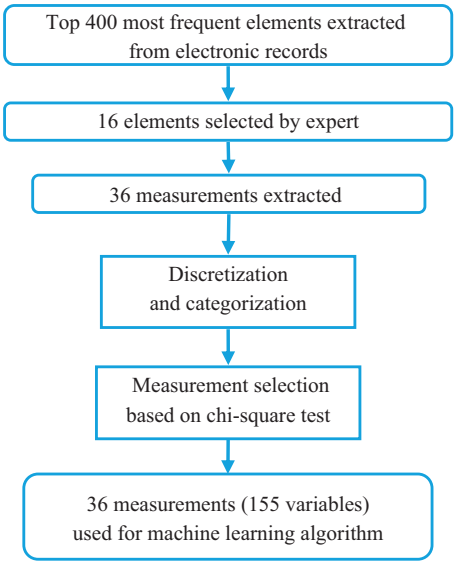

Fig. 2. Identification and selection procedure of clinical elements for machine learning algorithm.

Table 1
Predictive clinical elements.

All predictive clinical elements							
Clinical element	Type	Availability	Categorization				
Temperature	Continuous	924 (97.7%)	C_1 ≤36 °C	C_2 >36 °C and <38 °C	C_3 ≥38 °C		
Systolic blood pressure	Continuous	912 (96.4%)	Age group	C1	C2	C3	C4
			0–3 months	>60 and <80	≥80 or ≤60	≥100 or ≤50	≥130 or ≤45
			3–12 months	>80 and <100	≥100 or ≤80	≥120 or ≤70	≥150 or ≤60
			1–4 years	>90 and <110	≥110 or ≤90	≥125 or ≤75	≥160 or ≤65
			4–12 years	>90 and <120	≥120 or ≤90	≥140 or ≤80	≥170 or ≤70
			>12 years	>100 and <130	≥130 or ≤100	≥150 or ≤85	≥190 or ≤75
Oxygen saturation	Continuous	836 (88.4%)	C_1 >94	C_2 91–94	C_3 ≤90		
Heart rate	Continuous	930 (98.3%)	Age group	C_1	C_2	C_3	C_4
			0–3 months	>110 and <150	≥150 or ≤110	≥180 or ≤90	≥190 or ≤80
			3–12 months			≥170 or ≤80	≥180 or ≤70
			1–4 years	>100 and <150	≥150 or ≤100	≥150 or ≤70	≥170 or ≤60
			4–12 years			≥130 or ≤60	≥150 or ≤50
			>12 years	>90 and <120	≥120 or ≤90	≥120 or ≤50	≥140 or ≤40
				>70 and <110	≥110 or ≤70		
				>60 and <100	≥100 or ≤60		
Respiratory rate	Continuous	929 (98.2%)	Age group	C_1	C_2	C_3	C_4
			0–3 months	>29 and <61	≥61 or ≤29	≥81 or ≤19	≥91 or ≤15
			3–12 months	>24 and <51	≥51 or ≤24	≥71 or ≤19	≥81 or ≤15
			1–4 years	>19 and <41	≥41 or ≤19	≥61 or ≤15	≥71 or ≤12
			4–12 years	>19 and <31	≥31 or ≤19	≥41 or ≤14	≥51 or ≤10
			>12 years	>11 and <17	≥17 or ≤11	≥23 or ≤10	≥30 or ≤9
Level of consciousness	Narrative text	666 (70.4%)	C_1 Normal	C_2 Unclear	C_3 High risk		
Is patient experiencing pain?	Nominal	920 (97.2%)	C_1 Yes	C_2 No			
Cardiac	Nominal	881 (93.1%)	C_1 WNL	C_2 X			
Respiratory	Nominal	883 (93.3%)	C_1 WNL	C_2 X			
Neurologic	Nominal	875 (92.5%)	C_1 WNL	C_2 X			
Urinary	Nominal	872 (92.2%)	C_1 WNL	C_2 X			
Tissue perfusion and oxygenation	Nominal	858 (90.7%)	C_1 Excellent	C_2 Compromised	C_3 Adequate	C_4 Extremely compromised	
Development appropriate	Nominal	813 (85.9%)	C_1 Yes	C_2 No			
Acuity Level	Nominal	674 (71.2%)	C_1 Level 1	C_2 Level 2	C_3 Level 3	C_4 Level 4	C_5 Level 5
Work of breathing	Narrative text	758 (80.1%)	C_1 Normal	C_2 Unclear	C_3 High risk		
Perfusion cap refill	Nominal	398 (42.1%)	C_1 <3 s	C_2 ≥3 s			

	Level of consciousness	Work of breathing
Category definitions for the "level of consciousness" and "work of breathing" clinical elements		
Normal	Alert, active, quiet, arousable, WNL (within normal limits), responds, oriented, awake, answers questions, follows commands	Easy
Unclear	Sleeping, drowsy, emerging, jittery	Other comments
High risk	Lethargic, sedated, combative, unresponsive, seizure, agitated, seizing	Retracting, mild, nasal flaring, grunting, head bobbing, bobbing head, moderate, reports dyspnea, prolonged expiratory phase, gasping, severe

Note 1: The first five clinical elements were vital sign data entered into the electronic health record as numeric values by a nurse or nursing assistant. The remaining 10 clinical elements were entered by a nurse at least every 4 h as part of the body systems assessment. For this assessment the nurse chose from a list of pre-existing descriptors of the body system or selected "other" and input free text. *Note 2*: The second column in first section of table named "Availability" showed the total number and percentage of encounters having measurement for each clinical element in the training set; *Note 3*: C_i was used to indicate different categories under each element; *Note 4*: WNL and X denoted "Within Normal Limits" and "Exceptions to WNL"; *Note 5*: For "Acuity level", a locally developed indicator-based tool was used for assessing patient acuity and dependency regarding nurse resources.

controls. That is, to be conservative in our algorithm development, we used only a maximum of the first 23 hours' data points for the cases. Table 1 presents the details of the 16 clinical elements, which include five continuous, two narrative, and nine nominal elements.

In the fourth step, for each encounter, we extracted the most recent measurement in the study time-window (accounting for the 1 h cutoff threshold for cases) for each clinical element to develop a machine learning algorithm. For the five continuous clinical

Table 2
Performance comparison of logistic regression against Bedside PEWS and Monaghan's PEWS.

		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	AUC (95% CI)
Performance					
Training set	Bedside PEWS	0.715 (0.672–0.755)	0.708 (0.665–0.749)	0.710 (0.669–0.751)	0.785 (0.757–0.819)
	Monaghan's PEWS	0.741 (0.675–0.800)	0.830 (0.775–0.877)	0.797 (0.739–0.854)	0.814 (0.774–0.850)
	Logistic regression	0.827 (0.793–0.861)	0.873 (0.843–0.903)	0.867 (0.836–0.898)	0.919 (0.899–0.935)
Test Set	Bedside PEWS	0.736 (0.597–0.847)	0.717 (0.706–0.728)	0.021 (0.015–0.028)	0.816 (0.806–0.826)
	Monaghan's PEWS	0.684 (0.434–0.874)	0.816 (0.802–0.829)	0.023 (0.010–0.035)	0.744 (0.728–0.759)
	Logistic regression	0.849 (0.753–0.945)	0.859 (0.850–0.868)	0.048 (0.035–0.062)	0.912 (0.905–0.919)

		Sensitivity (improvement, *P*-value)	Specificity (improvement, *P*-value)	PPV (improvement, *P*-value)	AUC (improvement, *P*-value)
Performance comparison					
Training set	Logistic regression against Bedside PEWS	15.7%, $P<0.001$	23.3%, $P<0.001$	22.1%, $P<0.001$	17.1%, $P<0.001$
	Logistic regression against Monaghan's PEWS	11.6%, $P=0.01$	5.2%, $P=0.13$	8.8%, $P=0.03$	12.9%, $P<0.001$
Test set	Logistic regression against Bedside PEWS	15.4%, P = 0.15	19.8%, P < 0.001	125.4%, P < 0.001	11.8%, P < 0.001
	Logistic regression against Monaghan's PEWS	24.1%, P = 0.22	5.3%, P < 0.001	114.3%, P = 0.02	22.6%, P < 0.001

elements, we created four additional measurements, including the oldest, maximum, minimum and mean in the study time-window. With these four measurements, we intended to represent the dynamic nature of the patients' clinical conditions. In total, we collected 36 candidate measurements.

In the fifth step, we categorized each of these 36 measurements. For measurements from the nine nominal elements, we used the original labels to indicate the categories. For the five continuous elements, we performed a categorization step based on definitions of cut-off points identified from the published work[15] and guided by two physicians. We categorized the final two narrative elements based on keywords and synonyms provided by the physician. Table 1 shows the categorization of the 16 clinical elements. In our data set, the study variables rarely had missing values. Consequently, instead of attempting to impute values for the occasionally missing clinical elements, we added a new category "Not Available" ("N/A") for each clinical element and handled the "N/A" category as other naturally occurring categories of the data.[22] The categorization of these 36 measurements resulted in 155 dichotomous variables.

Finally, we used Chi-square calculations to test the significance of each measurement in the training set. All 36 measurements were selected and used to develop a machine learning algorithm at 0.05 *P*-value.

2.3. Algorithm development

We selected logistic regression as the Machine Learning (ML) algorithm and used Weka 3.6.8 as our experimental platform. We based the choice on logistic regression's wide usage in clinical decision systems and the relative ease of interpreting its output. In this study, instead of calculating a PEWS for each category, we used binary-valued variables to indicate the presence/absence of categories. We applied a forward stepwise approach with Akaike's Information Criterion (AIC) to select the best model.[23]

2.4. Evaluation metrics

To measure the algorithm's predictive performance, we calculated the sensitivity, specificity, positive predictive value (PPV) and AUC.[24] A predicted positive was any combination of predictor variables that had an output of >0.5 from the logistic regression model.

2.5. Experimental comparison

In order to evaluate the PICU transfer prediction methods, we implemented a standard 10-fold cross validation step. To assess the performance of the models on practical, real world data, we evaluated the models on the testing set, which represented the real world ratio of 24-h transfer and non-transfer encounters. For comparison, we also evaluated Parshuram's Bedside PEWS and Monaghan's PEWS.

When evaluating Parshuram's Bedside PEWS, we followed Robson's work[17] and included only five clinical elements available in our dataset instead of original seven (heart rate, systolic blood pressure, capillary refill time, respiratory rate, respiratory effort, transcutaneous oxygen saturation, and oxygen therapy). All of the elements except respiratory effort and oxygen therapy were also used in our machine learning algorithm.

Monaghan's PEWS is calculated for nearly all admitted patients at CCHMC. For the cases, we extracted the highest PEWS values recorded before the transfer event; for controls, we extracted the highest PEWS values recorded in the first 24 h of hospitalization. Four hundred twenty five encounters in the training set (consisting of 201 cases and 224 controls) included a non-missing value for Monaghan's PEWS and 3080 encounters in the test set (including 19 cases and 3061 controls) had Monaghan's PEWS. For a fair comparison, the prediction results for Monaghan's PEWS were calculated only on the encounters that had a non-missing Monaghan's PEWS value.

2.6. Timestamp experiment

In order to assess the robustness of our logistic regression model, we evaluated it using available clinical measurements at different points of the first 24-h after admission. Specifically, we ran the model 24 times using cumulative clinical measurements in each hour of the first 24-h after admission on the training set (i.e., 946 samples).

3. Results

3.1. Prediction results using baselines

Table 2 shows the results of Bedside PEWS at score 7 and Monaghan's PEWS at score 2. These cut-points were determined by

Table 3
Variables used in the final logistic regression model.

Variable	Odds ratio	*P*-value
Oldest temperature ≤36	0.78	0.222
Most recent temperature ≥38[b]	1.52	<0.001
Minimum temperature ≤36[b]	0.41	<0.001
Oldest systolic blood pressure C_4[a]	72.91	0.207
Most recent systolic blood pressure C_1[a]	0.67	0.070
Maximum systolic blood pressure C_1[a,b]	1.57	<0.001
Minimum systolic blood pressure C_4[a,b]	25.47	<0.001
Mean systolic blood pressure C_1[a,b]	0.47	<0.001
Most recent oxygen saturation >94[b]	0.43	<0.001
Maximum oxygen saturation >94	8.42	0.105
Minimum oxygen saturation ≤90[b]	1.80	0.021
Most recent heart rate C_1[a]	0.61	<0.001
Most recent heart rate C_3[a,b]	3.07	<0.001
Most recent heart rate C_4[a,b]	22.78	0.031
Maximum heart rate C_3[a,b]	0.58	<0.001
Minimum heart rate C_4[a,b]	24.22	0.002
Mean heart rate C_4[a,b]	2.68	0.039
Minimum respiratory rate C_4[a,b]	3.10	<0.001
Mean respiratory rate C_4[a,b]	4.55	<0.001
Level of consciousness High risk[b]	7.16	0.005
Is the patient experiencing pain? Yes[b]	1.97	<0.001
Cardiac within normal limits[b]	0.28	<0.001
Neurologic within normal limits[b]	0.26	<0.001
Tissue perfusion and oxygenation adequate[b]	89.04	<0.001
Tissue perfusion and oxygenation extremely compromised	28.65	0.246
Acuity level Level 3	1.52	0.611
Acuity level Level 4[b]	1.26	<0.001
Work of breathing High risk[b]	8.60	0.049
Perfusion cap refill <3 s[b]	2.10	0.031

[a] See Table 1.
[b] Indicates variables significantly associated with PICU transfer.

logistic regression with threshold 0.5 to guarantee a fair comparison.

3.2. Prediction results using logistic regression

We included 29 variables associated with 13 of the 16 clinical elements, in the final model by the forward stepwise approach (Table 3). Of the 29 variables, 23 were significantly associated with PICU transfer ($P < 0.05$). The results are presented in Table 2.

3.3. Results comparison

Table 2 presents the performance of the different predictive methods. In comparison to the Bedside PEWS and Monaghan's PEWS in the test set, our model's AUC and specificity were statistically significantly improved. Our model's sensitivity numerically improved by 15.4 and 24.1% in the test set, but these improvements were not statistically significant. We hypothesize that the relatively small number of cases in the test set hindered the detection of statistical significance. The PPV of our model was 8.1%, over twice the value of that found in our dataset of the two PEWS. As a previous study[25] described, the big decrease of PPV in the test set was caused by the domination of controls.

3.4. Timestamp experiment

Fig. 3 (top) shows the performance of our logistic regression model on different points (from 1 to 24 h) after admission. We can see that AUC increases significantly in the first 6 h and reaches the plateau after 7 h, which means that our model can perform a good prediction using the clinical measurements of first 7 h after admission. The bottom of Fig. 3 displays the PICU transfer distribution for the 473 cases in the training set. The average and median PICU transfer times are 11.7 and 11 h respectively, and 79% admissions are transferred to PICU after 7 h.

4. Discussion

We used existing clinical data in the EHR and machine learning to develop and validate a prediction algorithm for PICU transfer of hospitalized patients in the first 24 h. Through a process using expert clinician opinion, categorization and machine learning, we built a model consisting of 29 variables for predicting PICU transfer. Our algorithm achieved a 0.912 (95% CI 0.905–0.919) AUC in the test set. This result was statistically significantly higher than application of two existing PEWS in our test data set. Unlike previous PEWS which used a number of sub-scores to create an overall score with various cut-points, we used logistic regression so that the output was a percentage likelihood of PICU transfer. With this approach we were able to achieve 0.849 sensitivity and 0.859 specificity.

Our prediction algorithm performed significantly better than two published PEWS that were based on dynamic clinical elements, such as vital signs. One reason for this finding is that we used 29 variables from 16 clinical elements as compared to 3–7 variables in PEWS with which we compared. Our variables included vital signs, which both other scores employ. We also included level of consciousness, pain assessments, and work of breathing that each met two important criteria: (1) face validity in association with worsening patient status that might precede PICU transfer, and (2) were obtained by our nurses in the course of their usual clinical assessments. With the exception of one variable (capillary refill) each of our variables was available in >70% of encounters, with the majority being present in >90% of the encounters. At our center, these data did not require an extra reporting structure, additional clinical assessments, or research nurses. Each was present in the EHR for clinical care, but we believe each was poorly leveraged in the course of care in identifying and predicting patient risk.

The timestamp experiment showed that clinical measurements taken in the first 7 h were sufficient for our predictions. We found a relatively low PPV as transfer to the PICU in the first 24 h is an uncommon event. As we believe the cost of a false negative is considerably higher than a false positive, relatively low PPV may be a tolerable trade-off.

Our prediction algorithm can be integrated into our rapid response system to identify patients at elevated risk for PICU transfer. Current mechanisms to trigger or activate the rapid response system have limitations.[6] Early warning scores represent a quantitative and the most extensively validated mechanism to activate, but other activation mechanisms including subjective nurse, family, or physician concerns are also used. These subjective mechanisms have limitations regarding validity and reliability, and to date, PEWS are limited by the modest number of elements from which a score can be generated. Our method of machine learning/logistic regression allows an output of the percentage likelihood of PICU transfer to be calculated for an almost limitless number of clinical elements. While the best way to use this output will need to be determined prospectively, we believe a rapid response system could have multiple thresholds based on the percentage likelihood. For example, if the likelihood were >50% of PICU transfer within 24 h, this may prompt an automatic call of the medical emergency team for multidisciplinary assessment. A score of >25% might trigger a bedside evaluation by the primary medical team and a recalculation of prediction within 2 h. An output of >95% might put in motion, through clinical decision aids, a process that makes immediate PICU transfer the default action and a physician would need to take active action to avoid such a result.

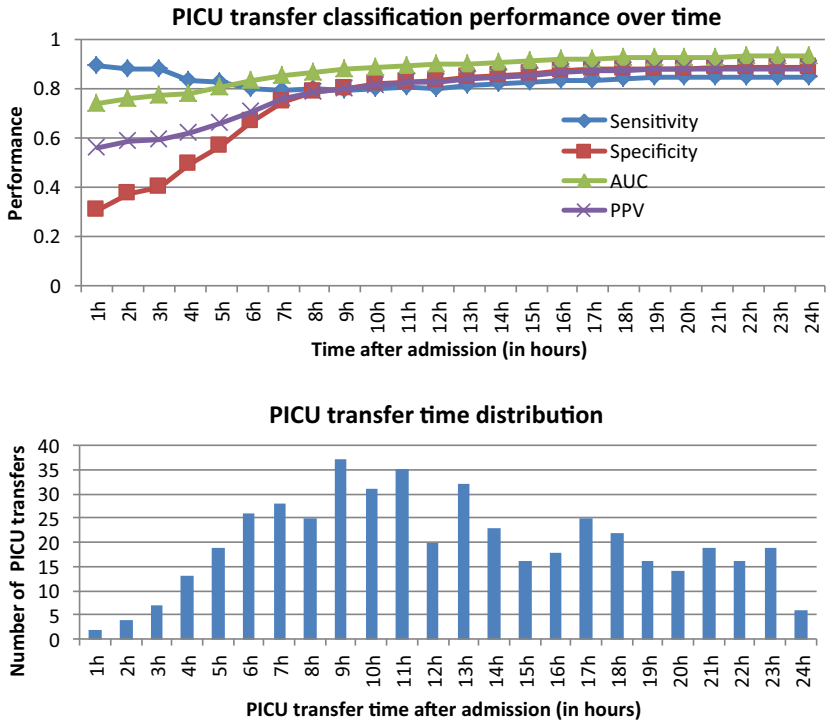

Fig. 3. Timestamp experiment results.

4.1. Limitations

Missing data was a major cause of incorrect prediction and we need to develop a proper imputation methodology. In the current study, we used a very simple method to address the challenges of missing data. We will implement more complex imputation methods in our future work. Imputation will be especially important because as we will add more variables to the model, the additional variables will include missing values more frequently.

Transfer to the PICU, while a clinically important event, does have some limitations as it may be driven in part by non-patient factors such as PICU bed availability. In future studies, instead of focusing exclusively on the need for PICU transfer as a dependent variable, we will predict the deterioration of hospitalized children's clinical status and will include other variables (e.g. calling a medical emergency team, nurse or physician identification as high risk) as dependent variables.

In this study, all the clinical elements had a much higher percentage of availability than in previous studies. The high percentage of available clinical measures provided the bases for applying machine learning to detect 24-h PICU transfer. This may reduce the generalizability of our findings in centers with more frequently missing data. In future studies, we will partner with other academic and non-academic children's hospitals to validate our algorithm in a diverse set of institutions on a prospective set of patients. Although our algorithm was created in the first 24 h, applying it after 24 h is quite straightforward. Similarly to the timestamp experiment, we need to regenerate the value of each variable every few hours (e.g., 1, 2, 4) and use the model to calculate the probability. However, the effectiveness of this approach needs to be verified on prospective data. The deterioration of hospitalized children is a complicated phenomenon with a variety of clinical antecedents and causes. A prolonged seizure that results in PICU transfer has a different pathophysiology and likely different clinical antecedents than worsening respiratory distress in pneumonia. In our current study, we only tested 16 clinical elements for prediction. We included just one static element (presence of developmental delay) and 15 dynamic elements. While we believe dynamic elements will likely be more useful in predicting PICU transfer in the short term, we suspect further addition of static elements such as a need for medical technology will improve prediction accuracy. We did not include lab test results, and an earlier work has explored the predictive potential of lab tests and medications for patient status deterioration.[26] In future works, we will include many more data points for the patients such as lab tests, medications, diagnostic history, social history, and family history.

5. Conclusion

On a retrospective data set we successfully developed a logistic regression algorithm that utilized the EHR content to predict PICU transfer for pediatric patients' first day of hospitalization. The novel algorithm achieved higher sensitivity, specificity and AUC than two of the current PEWS reported in the literature.

Author contributions

IS, PB and HZ designed the study. HZ ran the experiments, analyzed the results, created the tables and figures, and contributed to the draft and final manuscript. QL, TL, DW and YN contributed ideas for algorithm development. The project was supervised by IS. The first draft of the manuscript was prepared by HZ, IS, and PB with additional contributions by all authors. All authors read and approved the final manuscript.

Funding

The authors were supported by internal funds from Cincinnati Children's Hospital Medical Center. HZ, QL, TL, YN, and IS were partially supported by grants 5R00LM010227-05, 1R21HD072883-01, and 1U01HG006828-01.

Conflict of interest statement

No conflict of interest for any of the authors.

References

1. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. N Engl J Med 2009;361:1368–75.
2. Silber JH, Kaestner R, Even-Shoshan O, Wang Y, Bressler LJ. Aggressive treatment style and surgical outcomes. Health Serv Res 2010;45:1872–92.
3. Nadkarni VM, Larkin GL, Peberdy MA, et al. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. J Am Med Assoc 2006;295:50–7.
4. Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. Chest 1990;98:1388–92.
5. Jones DA, DeVita MA, Bellomo R. Rapid-response teams. N Engl J Med 2011;365:139–46.
6. DeVita MA, Smith GB, Adam SK, et al. "Identifying the hospitalised patient in crisis" – a consensus conference on the afferent limb of rapid response systems. Resuscitation 2010;81:375–82.
7. Bradman K, Borland M, Pascoe E. Predicting patient disposition in a paediatric emergency department. J Paediatr Child Health 2012, http://dx.doi.org/10.1111/jpc.12011.
8. Petrillo-Albarano T, Stockwell J, Leong T, Hebbar K. The use of a modified pediatric early warning score to assess stability of pediatric patients during transport. Pediatr Emerg Care 2012;28:878–82.
9. Sweney JS, Poss WB, Grissom CK, Keenan HT. Comparison of severity of illness scores to physician clinical judgment for potential use in pediatric critical care triage. Disaster Med Public Health Prep 2012;6:126–30.
10. Parshuram CS, Bayliss A, Reimer J, Middaugh K, Blanchard N. Implementing the Bedside Paediatric Early Warning System in a community hospital: a prospective observational study. Paediatr Child Health 2011;16:e18–22.
11. Skaletzky SM, Raszynski A, Totapally BR. Validation of a modified pediatric early warning system score: a retrospective case-control study. Clin Pediatr (Phila) 2012;51:431–5.
12. Parshuram CS, Duncan HP, Joffe AR, et al. Multicentre validation of the bedside paediatric early warning system score: a severity of illness score to detect evolving critical illness in hospitalised children. Crit Care 2011;15:R184.
13. Akre M, Finkelstein M, Erickson M, Liu M, Vanderbilt L, Billman G. Sensitivity of the pediatric early warning score to identify patient deterioration. Pediatrics 2010;125:e763–9.
14. Demmel KM, Williams L, Flesch L. Implementation of the pediatric early warning scoring system on a pediatric hematology/oncology unit. J Pediatr Oncol Nurs 2010;27:229–40.
15. Parshuram CS, Hutchison J, Middaugh K. Development and initial validation of the Bedside Paediatric Early Warning System score. Crit Care 2009;13:R135.
16. Tucker KM, Brewer TL, Baker RB, Demeritt B, Vossmeyer MT. Prospective evaluation of a pediatric inpatient early warning scoring system. J Spec Pediatr Nurs 2009;14:79–85.
17. Robson MA, Cooper CL, Medicus LA, Quintero MJ, Zuniga SA. Comparison of three acute care pediatric early warning scoring tools. J Spec Pediatr Nurs 2012, http://dx.doi.org/10.1016/j.pedn.2012.12.002.
18. Edwards ED, Powell CV, Mason BW, Oliver A. Prospective cohort study to test the predictability of the Cardiff and Vale paediatric early warning system. Arch Dis Child 2009;94:602–6.
19. Das S, Eisenberg LD, House JW, et al. Meaningful use of electronic health records in otolaryngology: recommendations from the American Academy of Otolaryngology – Head and Neck Surgery Medical Informatics Committee. Otolaryngol Head Neck Surg 2011;144:135–41.
20. Jha AK. The promise of electronic records around the corner or down the road? J Am Med Assoc 2011;306:880–1.
21. Sittig DF, Wright A, Osheroff JA, et al. Grand challenges in clinical decision support. J Biomed Inform 2008;41:387–92.
22. Grzymala-Busse W, Hu M. A comparison of several approaches to missing attribute values in data mining. In: Proceedings of the second international conference on rough sets and current trends in computing RSCTC'2000. 2000. p. 340–7.
23. Venables WN, Ripley BD. Modern applied statistics statistics with S-PLUS. 4h ed. New York: Springer; 2002.
24. Metz CE. Basic principles of ORC analysis. Semin Nucl Med 1978;8:283–98.
25. Lalkhen G, McCluskey A. Clinical tests: sensitivity and specificity. Crit Care Pain 2008;8:221–3.
26. Bonafide CP, Holmes JH, Nadkarni VM, Lin R, Landis JR, Keren R. Development of a score to predict clinical deterioration in hospitalized children. J Hosp Med 2012;7:345–9.