Early prediction of critical events for infants with single-ventricle physiology in critical care using routinely collected data

Check for updates

Victor M. Ruiz, MS,[a,b] Lucas Saenz, MD,[c] Alejandro Lopez-Magallon, MD,[d] Ashlee Shields, RN, MSN, CCRN,[c] Henry A. Ogoe, PhD,[e] Srinivasan Suresh, MD,[e] Ricardo Munoz, MD,[d] and Fuchiang R. Tsui, PhD[a,b,f,g]

ABSTRACT

Objective: Critical events are common and difficult to predict among infants with congenital heart disease and are associated with mortality and long-term sequelae. We aimed to achieve early prediction of critical events, that is, cardiopulmonary resuscitation, emergency endotracheal intubation, and extracorporeal membrane oxygenation in infants with single-ventricle physiology before second-stage surgery. We hypothesized that naïve Bayesian models learned from expert knowledge and clinical data can predict critical events early and accurately.

Methods: We collected 93 patients with single-ventricle physiology admitted to intensive care units in a single tertiary pediatric hospital between 2014 and 2017. Using knowledge elicited from experienced cardiac-intensive-care-unit providers and machine-learning techniques, we developed and evaluated the Cardiac-intensive-care Warning INdex (C-WIN) system, consisting of a set of naïve Bayesian models that leverage routinely collected data. We evaluated predictive performance using the area under the receiver operating characteristic curve, sensitivity, and specificity. We performed the evaluation at 5 different prediction horizons: 1, 2, 4, 6, and 8 hours before the onset of critical events.

Results: The area under the receiver operating characteristic curves of the C-WIN models ranged between 0.73 and 0.88 at different prediction horizons. At 1 hour before critical events, C-WIN was able to detect events with an area under the receiver operating characteristic curve of 0.88 (95% confidence interval, 0.84-0.92) and a sensitivity of 84% at the 81% specificity level.

Conclusions: Predictive models may enhance clinicians' ability to identify infants with single-ventricle physiology at high risk of critical events. Early prediction of critical events may indicate the need to perform timely interventions, potentially reducing morbidity, mortality, and health care costs. (J Thorac Cardiovasc Surg 2019;158:234-43)
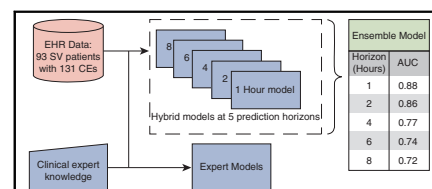
Development and evaluation of models for prediction of CEs in infants with SV physiology.

Central Message

C-WIN models achieved early prediction of critical events in infants with single-ventricle physiology by leveraging routinely collected data, artificial intelligence, and expert clinical knowledge.

Perspective

Infants with SV physiology are at high risk of unexpected deterioration before stage 2 palliation. We developed the C-WIN system using routinely collected clinical data, machine learning, and clinical knowledge. C-WIN predicted critical events with up to 8 hours of anticipation and may help clinicians perform early interventions to reduce morbidity, mortality, and health care costs.

Children and adults with congenital heart disease are a diverse and complex population whose management presents a challenge for clinicians and a heavy burden for health care systems. In the United States, 3.7% of pediatric hospitalizations and 15% of the total cost of pediatric hospitalizations are related to congenital heart disease.[1]

Abbreviations and Acronyms
AUC   = area under the curve
BP   = blood pressure
CE   = critical event
CPR   = cardiopulmonary resuscitation
CPT   = conditional probability table
C-WIN = Cardiac-intensive-care Warning INdex
ECMO = extracorporeal membrane oxygenation
EEI   = emergency endotracheal intubation
EHR   = electronic health record
ICU   = intensive care unit
NB   = naïve Bayes
SV   = single ventricle

Scanning this QR code will take you to the article title page to access supplementary information.

Infants with single-ventricle (SV) physiology are among the most complex congenital heart disease subpopulations and have high mortality and morbidity risk before stage 2 surgical repair.[1] Although in-hospital mortality has decreased,[2-4] patient deterioration may happen unexpectedly during the course of critical care. Even under the care of experienced critical-care teams, potentially catastrophic critical events (CEs), such as cardiopulmonary resuscitation (CPR), emergency endotracheal intubation (EEI), and extracorporeal membrane oxygenation (ECMO), are common in patients with SV physiology. Such events may negatively affect morbidity, mortality, and hospital length of stay.[5,6]

Early warning systems are clinical decision support aids that can help clinicians detect patient deterioration by assessing patients' risk of CEs in real time and providing alerts. To be effective, these systems should provide accurate and meaningful alerts without contributing to increased clinician workload or alert fatigue.[7,8]

Models that predict mortality risk in general intensive care unit (ICU) populations have been available for decades.[9,10] However, models that predict CEs in specialized populations such as infants with SV physiology are scarce. Although several early warning systems for CEs have been validated in non-ICU[11-13] and ICU[14-18] populations, only a few have been validated in our target population.[19-21] To the best of our knowledge, Rusin and colleagues[21] published the only model for infants with SV physiology in the ICU that can predict CEs in real-time. This model exhibited high accuracy in the hour

preceding CEs. Nonetheless, this state-of-the-art model has limitations. First, its accuracy decreases rapidly when CEs are predicted more than 2 hours in advance. Second, it relies on real-time analysis of electrocardiogram waveforms sampled at 240 Hz and vital signs sampled at 0.5 Hz, which may present technologic and financial challenges for many institutions.

To address the need of an early warning system for the SV ICU population, we developed the Cardiac-intensive-care Warning INdex (C-WIN) system. C-WIN is composed of predictive models that leverage routinely collected data for the prediction of CEs (CPR, EEI, and ECMO) in infants with SV physiology in pediatric ICUs, including pediatric ICU, neonatal ICU, and cardiac ICU. We built these models through the elicitation of clinical knowledge from clinical experts and the application of machine-learning techniques.

We hypothesized that C-WIN can accurately predict the onset of CEs from routinely collected data with several hours of anticipation and that machine-learning techniques can significantly improve the performance of expert-based models. The contributions of this study are 2-fold. First, we developed models that incorporated expert knowledge and applied machine-learning techniques to achieve state-of-the-art early prediction of CEs. Second, we analyzed how the predictive performance of our models changed on the basis of the lead time before CEs used to select data for model training.

MATERIALS AND METHODS
In this section, we describe the target cohort, definition of cases and controls, data processing, model development, and performance evaluation. The Institutional Review Board at the University of Pittsburgh approved this study (PRO17020157).

Data Collection
We retrieved inpatient data of infants admitted to the ICUs (pediatric ICU, neonatal ICU, and cardiac ICU) at the University of Pittsburgh Medical Center Children's Hospital of Pittsburgh. We included hospital admissions of infants less than 6 months old who had SV diagnoses and were admitted to ICUs between January 2014 and August 2017. We excluded hospital admissions of patients who had undergone second surgical palliation (bidirectional Glenn). Figure 1 describes the data retrieval process, and Table E1 in the Supplemental Material lists the International Classification of Diseases codes used to identify those with SV physiology.

Case Definition
We defined cases as CEs that occurred in the ICU before patients underwent stage 2 surgery and at least 8 hours after patients' first ICU admission during a hospital admission. We considered multiple CEs from the same patient as independent cases if they occurred at least 8 hours after previous CEs. This allowed us to compare predictive performance at different prediction horizons (time before CEs) on the same cases with up to 8 hours of anticipation.

We used documented times of clinical events available in the Cerner electronic health record (EHR) system at the University of Pittsburgh Medical Center Children's Hospital of Pittsburgh as the gold standard to identify the start and end times of CEs. For EEI, the start time (intubation) was the time when airway changes from *natural* to *artificial* were documented
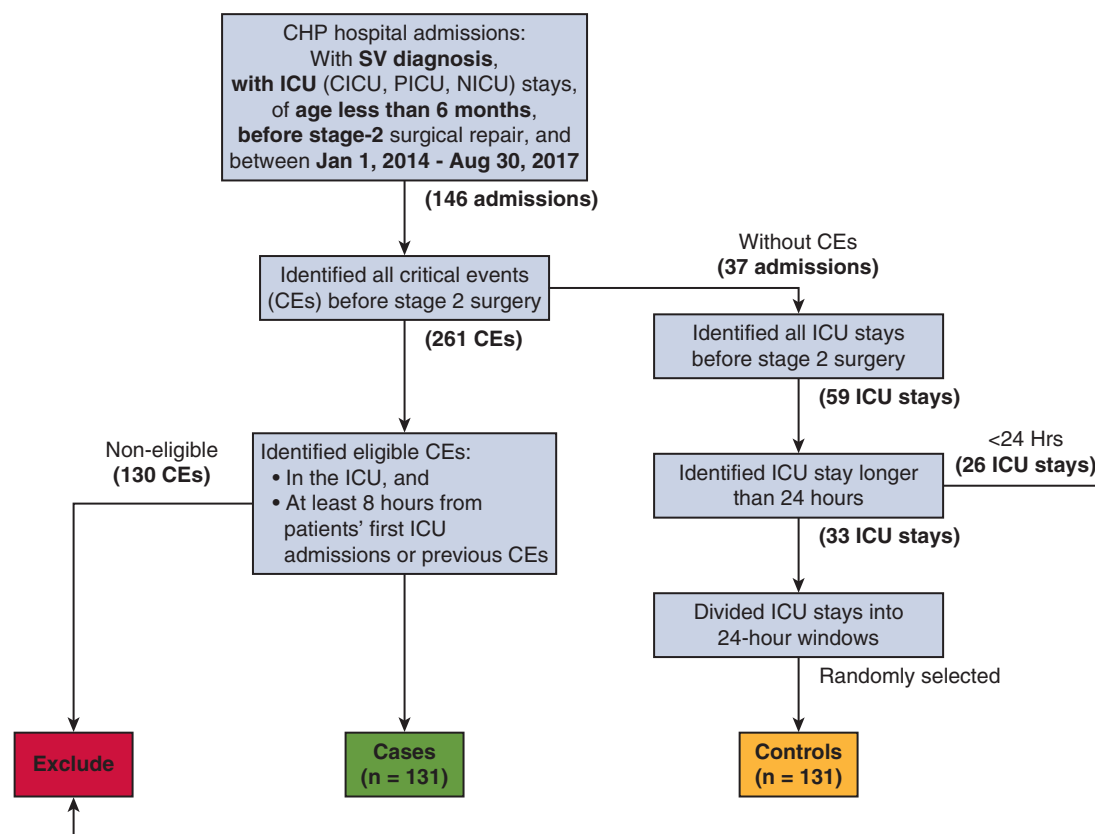
CONG

FIGURE 1. Data retrieval and case-control definition process; multiple CEs during the same hospital admission were considered as different cases and were eligible if they occurred at least 8 hours after patients' first ICU admission or a previous CE. ICU stays from patients who did not experience CEs during their hospital admission and were longer than 24 hours were selected as controls. Control ICU stays were divided into 24-hour windows, and all windows were considered as separate controls. We randomly selected 131 control windows to match the number of CE cases. *CHP*, University of Pittsburgh Medical Center Children's Hospital of Pittsburgh; *SV*, single ventricle; *ICU*, intensive care unit; *CICU*, cardiac intensive care unit; *PICU*, pediatric intensive care unit; *NICU*, neonatal intensive care unit; *CE*, critical event.

in nursing and respiratory therapy charts, and the end time (extubation) was the time when airway changes from *artificial* to *natural* were documented by a respiratory therapist. For ECMO, start and end times were the times of cannulation and decannulation, respectively, which were documented by perfusionists in charge of managing ECMO circuits. For CPR, start times were the time when specific keywords were documented in nursing communication notes. Keywords included *arrest, arrest sheet, arrest code, chest compressions, condition A,* and *CPR.*

Control Definition

We defined controls as periods of ICU stays longer than 24 hours from patients who did not experience CEs during their hospital admission. We divided ICU stays into 24-hour windows and used each window as a distinct control. Figure 1 describes the case-control definition process.

Expert Model Development

We built expert-based models using knowledge elicited from expert clinicians (1 pediatric critical-care cardiologist and 1 pediatric cardiologist, assisted by 2 cardiac ICU nurse specialists). Expert knowledge included a list of 34 routinely collected clinical variables that experts identified as relevant for predicting CEs, meaningful numeric ranges for each variable, and contingency tables that quantified experts' assessment of the importance of each variable. We mapped expert contingency tables to discrete conditional probability tables (CPTs),

which denote the extent to which observing a variable value influences the probability of experiencing CEs according to expert opinion. Variables included vital signs and laboratory-test results (a full list is available in Table E2 in the Supplemental Material). Experts also identified a subset of 16 variables considered the minimal set of relevant variables.

We used expert CPTs to parametrize naïve Bayes (NB) network models. These models assume that predictors are conditionally independent given the value of the outcome variable (CEs) as indicated by the following equation:

$$P(X_1, X_2, \ldots, X_n | CE) = \prod_{i=1}^{n} P(X_i | CE)$$

where X_i is a predictor (eg, blood pressure [BP]) and *CE* is the outcome variable. Bayes' theorem is used to compute the posterior probability of CEs given observed predictor values, as shown in the following equation:

$$P(CE | X_1, X_2, \ldots, X_n) = \frac{\prod_{i=1}^{n} P(X_i | CE) P(CE)}{\prod_{i=1}^{n} P(X_i | CE) P(CE) + \prod_{i=1}^{n} P(X_i | \overline{CE}) P(\overline{CE})}$$

where X_i is the *i*-th predictor, *n* is the number of predictors, and $P(\overline{CE}) = 1 - P(CE)$.

We built 4 NB models based exclusively on elicited knowledge (2 from each expert). The NB-expert1-full and NB-expert2-full models used all 34 available variables, and the NB-expert1-lean and NB-expert2-lean models
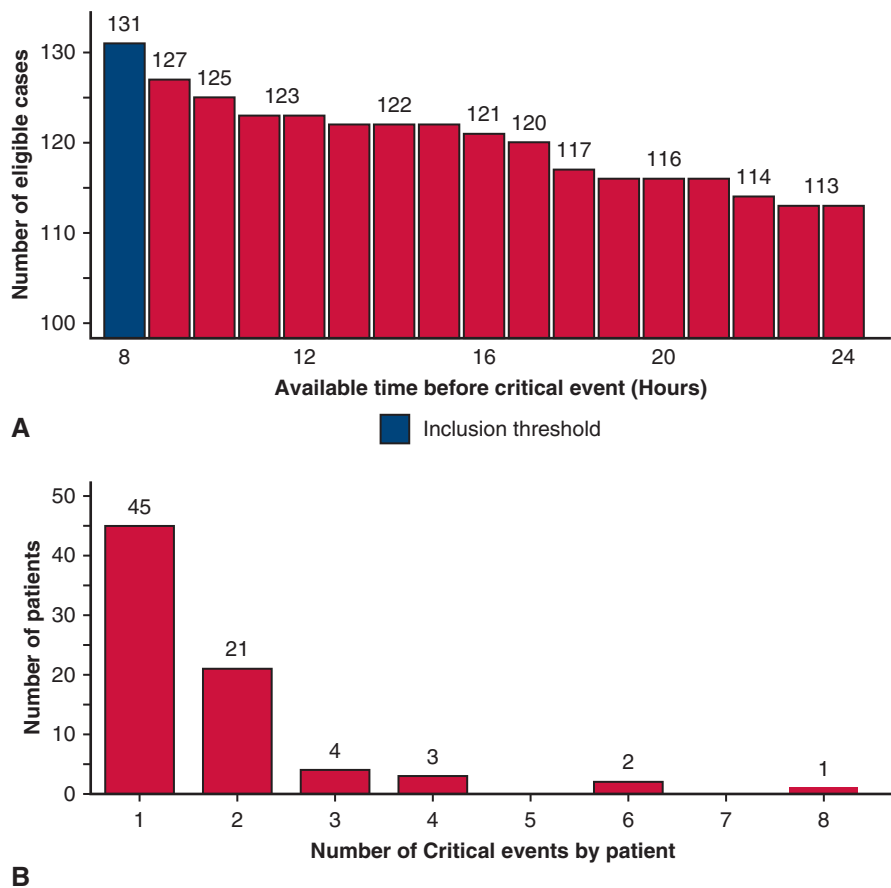
FIGURE 2. Distribution of CEs by data availability and patient. A, Number of cases (CEs) eligible for inclusion depending on data availability, that is, the time between a patient's first ICU admission or previous CEs and the onset of a CE. B, Distribution of the number of eligible CEs by patient. Eligible events were those that occurred in the ICU and at least 8 hours after patients' first ICU admission or previous CEs (131 in total).

used the minimal subset of 16 variables. Each expert defined his/her own variable ranges and CPTs.

Machine-Learning Model Development

We used clinical data to build additional NB models that combined expert knowledge and machine-learning techniques. These models,

TABLE 1. Distribution of primary diagnosis among patients with single-ventricle physiology included in the study population

Diagnosis	No. of patients (%)
Hypoplastic left heart syndrome	40 (43)
Congenital mitral stenosis	18 (19.4)
Pulmonary valve atresia	12 (12.9)
Hypoplastic right heart syndrome	6 (6.5)
Congenital atresia and stenosis of aorta	5 (5.4)
Tricuspid atresia and stenosis, congenital	5 (5.4)
Common ventricle	3 (3.2)
Congenital stenosis of aortic valve	2 (2.2)
Double inlet left ventricle	2 (2.2)
Total	93 (100)

identified by the NB-ML prefix, differed from NB-expert models in 2 ways. First, they used all 34 variables (NB-ML-full models) or a subset of variables selected based on information gain scores[22] (NB-ML-lean models) with an empirical threshold of 0.01. Information gain ranges between 0 and 1, denoting no prediction ability or perfect prediction, respectively. Second, NB-ML models' CPTs were computed with maximum-likelihood estimates obtained from clinical data.

We estimated CPTs from variable values (patient states) at different prediction horizons. Thus, CPTs estimated from patient states at 8 hours before CEs may differ substantially from those estimated from patient states at 2 hours before CEs. We built NB-ML models using patient states at 1, 2, 4, 6, and 8 hours before CEs, as described in the Graphical Abstract (available online). For example, the NB-ML-lean-2 model was built with variables selected using information gain scores and CPTs estimated from patient states at 2 hours before CEs. We expected models built based on patient states at a given prediction horizon to achieve the best performance when predicting CEs with the same lead time. Finally, we built 2 ensemble models (NB-ML-avg-full and NB-ML-avg-lean) whose predictions were the average of the predictions from either all NB-ML-full or all NB-ML-lean models.

Data Processing

We processed longitudinal values of each variable in 3 steps: variable-value priority selection, uniform-window resampling, and missing-value imputation as described next.

CONG

Variable-Value Priority Selection

Model variables can be coded with different clinical concepts in an EHR system. For example, diastolic BP may be represented by *arterial diastolic pressure* (invasive measurement) and *diastolic BP* (noninvasive). For each variable, we identified all relevant EHR clinical concepts and aggregated their measurements into a single series of values in chronologic order. We defined priority rules to select variable values when multiple clinical concepts were available simultaneously (eg, choose invasive over noninvasive BP values if both are available). Table E2 in the Supplemental Material shows the clinical concepts associated to each variable and their priority levels.

Uniform-Window Variable Resampling

We resampled variable values every 30 minutes, ending at the time of CEs for cases and at the time of discharge for controls. We averaged values of the same variable available during the same 30-minute window.

Missing Value Imputation

Although some variables (eg, BP, heart rate) are measured hourly, others (eg, lactate, base excess) may be measured at irregular intervals of several hours. Moreover, some variables (eg, a disintegrin and metalloproteinase with thrombospondin motifs, also known as "ADAMTS 13 activity") are measured rarely, leading to missing data. We imputed missing values with previous observations within 6 hours or with the variable's mean value in the training dataset if no previous measurements were available.

Prediction Performance Evaluation

We evaluated predictive performance in 10-fold cross-validation[23] with 3 metrics, namely, the area under the curve (AUC), sensitivity, and specificity. AUC values of 0.5 and 1 denote no predictive ability and perfect prediction, respectively. We evaluated each model at 1, 2, 4, 6, and 8 hours before the onset of CEs, thus assessing whether performance deteriorated as we predicted CEs further into the future. We used 2-sided DeLong tests[24] with Bonferroni multiple-hypotheses correction[25] to compare AUCs.

We used 2 criteria to select the values used to generate predictions (posterior probabilities of CEs). For cases, we selected variable values known during the 30-minute window ending at the time of presentation of CEs. For controls, we selected variable values from the mid-point (12th hour) of each 24-hour control period.

Secondary Analyses

We tested whether assumptions in the inclusion and exclusion criteria of CEs affected model performance. First, we relaxed the CE inclusion criteria by including CEs that occurred at least 2 hours (instead of 8 hours) after first ICU admission or previous CEs and measured the performance of the NB-ML-full-2 model. Second, we applied a stricter CE inclusion criteria and measured the performance of the NB-ML-avg-model keeping only 1 CE per hospital admission in the test dataset. We made no changes to the models or to the set of controls in the test dataset for these analyses.

RESULTS

Cohort Description

We identified 146 hospital admissions in the study period. These corresponded to 120 patients, 223 ICU stays, and 261 CEs. A total of 93 patients and 131 CEs met our inclusion criteria for analysis, including 118 EEI, 9 ECMO, and 4 CPR events. Six of the included EEIs preceded further patient deterioration. Specifically, 5 preceded ECMO (3, 11, 21, 652, and 929 hours after intubation) and 1 preceded CPR (30 minutes after intubation). We randomly selected 131 control periods to match the number of cases.

TABLE 2. Classification of patients with single-ventricle physiology in the study population by type of palliative surgical procedure

Type of procedure	No. of patients (%)
Norwood	27 (29)
Modified Blalock–Taussig shunt	15 (16.1)
Other	15 (16.1)
Pulmonary artery banding	9 (9.7)
Hybrid	7 (7.5)
Bidirectional Glenn	2 (2.2)
Nonsurgical	18 (19.4)
Total	93 (100)

Patients in the nonsurgical category met the inclusion criteria for analysis but did not undergo palliative procedures during the study period. Two patients underwent bidirectional Glenn as their first palliative procedure. However, only ICU admissions before patients' stage 2 repair were used for model development and evaluation. Procedures in the "other" category included aortic arch repair, tetralogy of Fallot repair after Blalock–Taussig shunt, right ventricle to pulmonary artery conduit repair, aortopexy, valvuloplasty, stent placement in catheterization laboratory, atrioventricular septal defect repair after pulmonary artery banding and aortic arch repair, and Nikaidoh after Blalock–Taussig shunt.

Although most patients had only 1 eligible CE during their ICU stays, some patients experienced up to 8. Figure 2 shows the distribution of eligible CEs per patient and the number of eligible CEs depending on the amount of available data.

Hypoplastic left heart syndrome was the most common primary diagnosis (43% of study population) followed by

Time of critical event relative to surgical procedure

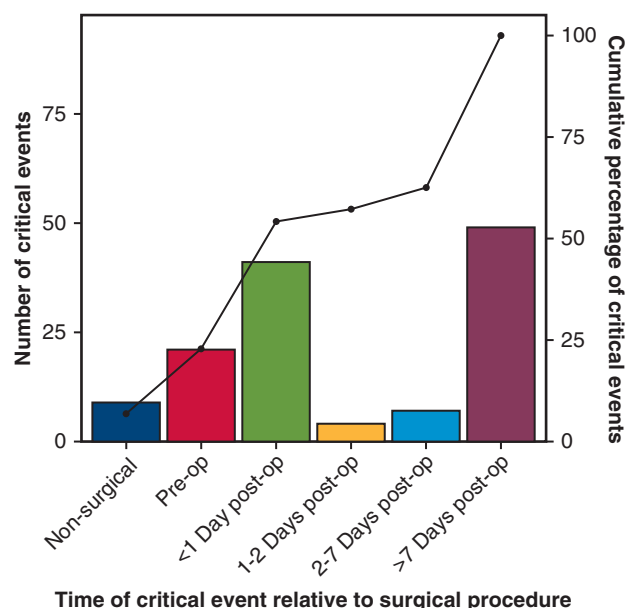**FIGURE 3.** Time of presentation of CEs relative to the time of palliative procedures; CEs included 131 EEIs, ECMO cannulations, and CPRs experienced by infants with SV physiology before stage 2 palliative surgery. The nonsurgical category corresponds to CEs experienced by patients who did not undergo palliative procedures during the study period. The *black line* shows a cumulative percentage of CEs.

TABLE 3. Prediction performance of naïve Bayes models at different prediction horizons

Approach	Prediction horizon	Model			
		NB-expert1-full (34 variables)	NB-expert2-full (34 variables)	NB-expert1-lean (16 variables)	NB-expert2-lean (16 variables)
Expert knowledge	1 h	**0.68 (0.61-0.75)**	0.58 (0.51-0.65)	0.6 (0.53-0.67)	0.5 (0.43-0.58)
	2 h	**0.71 (0.64-0.77)**	0.61 (0.54-0.68)	0.64 (0.57-0.71)	0.54 (0.46-0.61)
	4 h	**0.58 (0.51-0.65)**	0.57 (0.5-0.64)	0.48 (0.41-0.56)	0.5 (0.43-0.57)
	6 h	0.52 (0.45-0.59)	**0.54 (0.47-0.61)**	0.45 (0.38-0.52)	0.49 (0.42-0.56)
	8 h	**0.49 (0.42-0.56)**	0.46 (0.39-0.54)	0.41 (0.34-0.48)	0.42 (0.35-0.49)
		NB-ML-full-1 (34 variables)	NB-ML-full-2 (34 variables)	NB-ML-full-4 (34 variables)	NB-ML-full-6 (34 variables)	NB-ML-full-8 (34 variables)
Expert knowledge + data derived CPTs	1 h	**0.88 (0.84-0.92)**	0.88 (0.83-0.92)	0.84 (0.79-0.89)	0.79 (0.74-0.84)	0.72 (0.66-0.78)
	2 h	0.87 (0.82-0.91)	**0.87 (0.83-0.91)**	0.84 (0.79-0.89)	0.78 (0.72-0.84)	0.69 (0.63-0.76)
	4 h	0.78 (0.72-0.83)	0.77 (0.71-0.82)	**0.79 (0.74-0.85)**	0.75 (0.69-0.81)	0.71 (0.64-0.77)
	6 h	0.73 (0.66-0.79)	0.73 (0.66-0.79)	0.76 (0.7-0.81)	**0.74 (0.68-0.8)**	0.73 (0.67-0.8)
	8 h	0.66 (0.59-0.72)	0.68 (0.61-0.74)	0.72 (0.65-0.78)	0.72 (0.66-0.78)	**0.73 (0.67-0.79)**
		NB-ML-lean-1 (23 variables)	NB-ML-lean-2 (24 variables)	NB-ML-lean-4 (18 variables)	NB-ML-lean-6 (23 variables)	NB-ML-lean-8 (19 variables)
Expert knowledge + data derived CPTs + feature selection	1 h	0.87 (0.83-0.92)	**0.88 (0.83-0.92)**	0.83 (0.78-0.88)	0.77 (0.71-0.82)	0.69 (0.62-0.76)
	2 h	0.86 (0.81-0.91)	**0.87 (0.82-0.91)**	0.84 (0.78-0.89)	0.78 (0.72-0.83)	0.67 (0.6-0.74)
	4 h	0.77 (0.72-0.83)	0.77 (0.71-0.82)	**0.8 (0.74-0.85)**	0.74 (0.68-0.8)	0.69 (0.62-0.75)
	6 h	0.71 (0.64-0.77)	0.72 (0.66-0.78)	**0.75 (0.69-0.81)**	0.74 (0.67-0.8)	0.73 (0.67-0.79)
	8 h	0.64 (0.57-0.71)	0.67 (0.6-0.74)	0.71 (0.65-0.77)	0.72 (0.66-0.79)	**0.74 (0.68-0.8)**
		NB-ML-avg-full (34 variables)			NB-ML-avg-lean (24 variables)	
Ensemble	1 h	**0.88 (0.83-0.92)**			0.86 (0.82-0.91)	
	2 h	**0.86 (0.81-0.9)**			0.85 (0.8-0.9)	
	4 h	**0.77 (0.72-0.83)**			0.77 (0.71-0.83)	
	6 h	0.74 (0.68-0.8)			**0.75 (0.69-0.8)**	
	8 h	**0.72 (0.65-0.78)**			0.71 (0.65-0.77)	

Values are areas under the receiver operating characteristics AUC. Values in parentheses show 95% confidence intervals computed with 2000 bootstrap replicates. Boldface values show the best AUCs for each prediction horizon, that is, the number of hours before CEs when predictions were generated. The variables used in lean models were selected on the basis of expert opinion for the NB-expert1-lean and NB-expert2-lean models and based on information gain scores for NB-ML-lean models. *NB*, Naïve Bayes; *ML*, machine learning; *CPT*, conditional probability table.

mitral stenosis (19.4%), as shown in Table 1. Norwood was the most common palliative procedure (29% of study population) followed by Blalock–Taussig shunt (16.1%), as shown in Table 2. The majority (37.4%) of included CEs occurred 7 or more days after patients underwent palliative procedures. Figure 3 shows the distribution of times when CEs occurred relative to palliative procedures.

In our secondary analyses, we identified 154 CEs that happened at least 2 hours after first ICU admission or previous CEs. Additional CEs not included in the primary analysis included 21 EEI events, 1 ECMO event, and 1 CPR event.

Prediction Performance

Table 3 shows the AUC of all of our 16 models for all prediction horizons. Among purely expert-based models, the NB-expert1-full model achieved the best performance with AUCs ranging from 0.49 to 0.71. Models that combined expert knowledge and machine-learning techniques (NB-ML-full models) achieved statistically significantly higher AUCs compared with expert models for every prediction horizon (Bonferroni-adjusted $P < .01$). They achieved AUCs of 0.88 (NB-ML-full-1 model) and 0.73 (NB-ML-full-8 model) 1 and 8 hours before CEs, respectively. Receiver operating characteristic analysis showed that at 1 hour before CEs, the NB-ML-full-1 model had a sensitivity of 84% at the 81% specificity level and a sensitivity of 49% at the 95% specificity level. At 8 hours before CEs, the NB-ML-full-8 model had a sensitivity of 57% at the 80% specificity level or a sensitivity of 25% at the 95% specificity level.

The ensemble model with feature selection (NB-ML-avg-lean) achieved similar performance to that of the full ensemble model and reduced the variable space from 34 to 24. Among selected features, inspired oxygen fraction, calcium, base excess, mixed venous saturation, glucose, carbon dioxide, creatinine change from baseline,

hematocrit, lactate, and creatinine had the highest information gain scores at 1 hour before CEs.

The AUC of our predictive models increased as the prediction horizon approached the time of CEs. Figure 4 shows how the NB-expert1-full, NB-ML-avg-full, and every NB-ML-full model performed at each prediction horizon. We observed that the NB-ML-avg-full model's performance deteriorated less over time compared with all other NB-ML-full models. The NB-ML-avg-full and NB-ML-lean ensemble models predicted CEs by averaging the predictions produced by all NB-ML-full or NB-ML-lean models, respectively. Figure 5 shows the receiver operating characteristics curves of the NB-ML-avg-full model at every prediction horizon and sensitivity and specificity values for possible prediction thresholds.

In our secondary analyses, the NB-ML-full-2 had an AUC of 0.84 (0.79-0.89) at 2 hours before CEs after adding CEs that happened between 2 and 8 hours before first ICU admission or previous CEs to the test dataset. Finally, when evaluated on a test set with 1 CE per hospital admission (74 EEIs, 4 ECMOs, and 3 CPRs), the NB-ML-avg-full model had AUCs of 0.88 (0.82-0.94) and 0.76 (0.67-0.83) at 1 and 8 hours before CEs, respectively.

DISCUSSION

Hospital care of infants with SV physiology is complex because of their unique physiology, elevated severity of illness, and unpredictable clinical deterioration. In this study, we developed the C-WIN models, which achieved early prediction of CEs in this population. We addressed this need by developing predictive models that use objective and routinely collected data that can be retrieved automatically from an EHR system. Rather than following a completely data-driven approach, we incorporated expert knowledge into our models, which we believe is important to facilitate the adoption of predictive models into clinical workflows (Video 1).

We tested the hypothesis that models built from expert knowledge could predict CEs. We found that expert models achieved modest performance from 1 to 4 hours before CEs. However, they lost predictive power after 6 hours before CEs. A possible explanation for this behavior is that the CPTs provided by clinicians reflect the characteristics of patients close to the time of decompensation.

We also tested the hypothesis that applying machine-learning techniques could increase performance significantly. We found that computing CPTs from retrieved training data resulted in statistically significantly higher performance compared with using expert-defined CPTs. Furthermore, using information gain reduced model complexity. With a conservative selection threshold, the NB-ML-lean models achieved similar performance than that of NB-ML-full models using fewer variables.
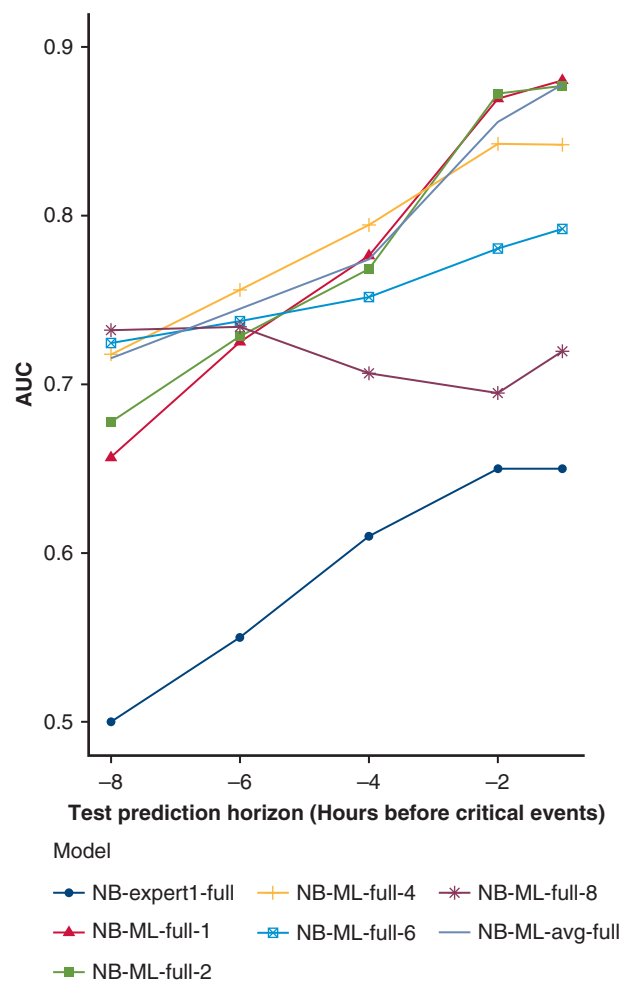

FIGURE 4. AUC over time of NB-Expert1, NB-ML-avg-full, and NB-ML-full models; curves show how each model performed at different prediction horizons, thus describing the difference between early prediction performance and prediction performance close to the time of critical events (CEs). The models used 34 variables identified by pediatric cardiologists as relevant for the prediction of CEs. They have NB structure and differ in the method used to compute their conditional probability tables (CPTs). The NB-Expert1 model was built exclusively from expert knowledge. Alternatively, the NB-ML-full models used CPTs derived from clinical data at different times before the onset of CEs, as denoted by their numeric suffix (eg, the NB-ML-full-2 model used patient states 2 hours before CEs). The NB-ML-avg-full model is an ensemble classifier whose predictions are the mean value of predictions from all NB-ML-full models. *AUC*, Area under the curve; *NB*, naïve Bayes; *ML*, machine learning.

Although NB networks are relatively simple compared with other modeling strategies (eg, deep artificial neural networks, support vector machines), we believe that they are an appropriate choice for this work. They have been used in biomedical research since the 1960s and are well suited for clinical applications.[26,27] Their computation is efficient, they can operate with missing and categoric data explicitly, and they can incorporate prior knowledge from clinicians or clinical data. Moreover, their simple
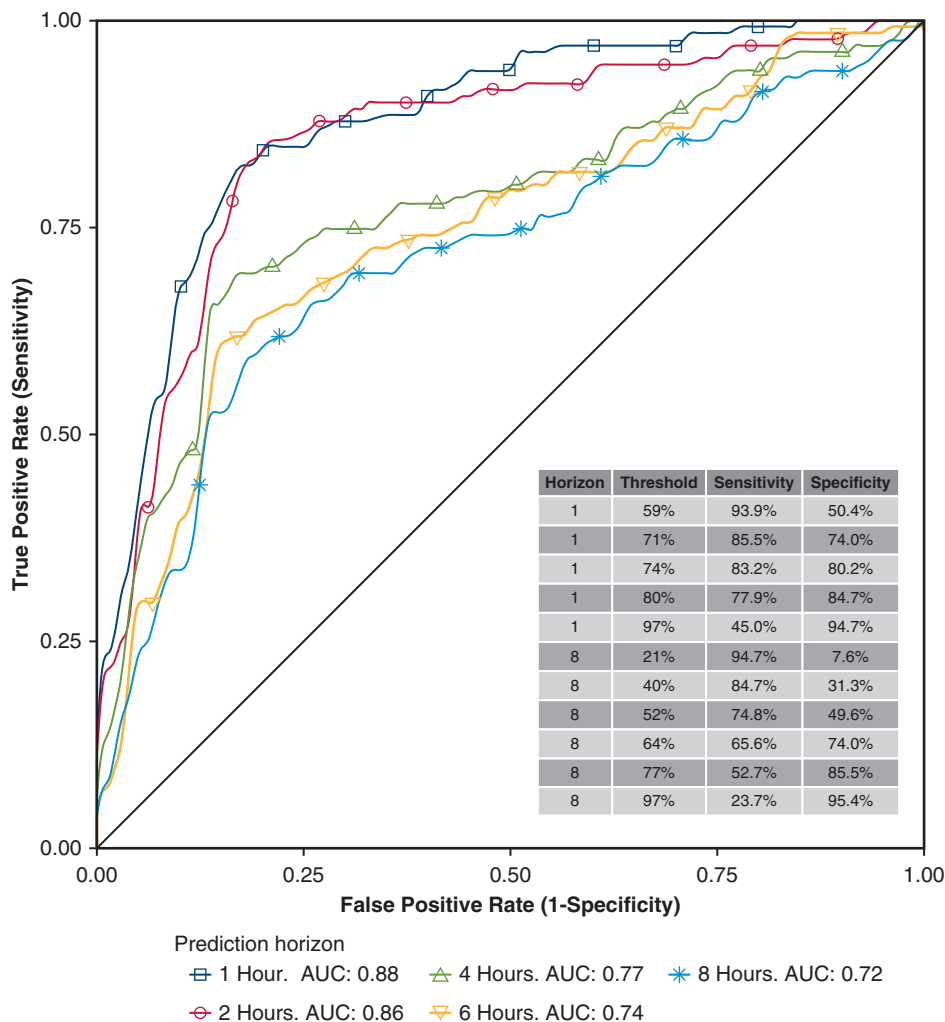
Horizon	Threshold	Sensitivity	Specificity
1	59%	93.9%	50.4%
1	71%	85.5%	74.0%
1	74%	83.2%	80.2%
1	80%	77.9%	84.7%
1	97%	45.0%	94.7%
8	21%	94.7%	7.6%
8	40%	84.7%	31.3%
8	52%	74.8%	49.6%
8	64%	65.6%	74.0%
8	77%	52.7%	85.5%
8	97%	23.7%	95.4%

Prediction horizon
- 1 Hour. AUC: 0.88
- 2 Hours. AUC: 0.86
- 4 Hours. AUC: 0.77
- 6 Hours. AUC: 0.74
- 8 Hours. AUC: 0.72

FIGURE 5. Receiver operating characteristics of the NB-ML-avg-full model. This model predicted critical events (CEs) by averaging the predictions of all NB-ML-full models, which were built based on 34 clinical variables identified by pediatric cardiologists as relevant for the prediction of CEs. The conditional probability tables (CPTs) of variables in NB-ML-full models were computed on the basis of maximum-likelihood estimates obtained from clinical data at different times before the onset of CEs. The table in the lower-right corner shows specificity and sensitivity values of the NB-ML-avg-full model at 1 and 8 hours before CEs for possible selections of prediction threshold, that is, the minimum value of the model's prediction (posterior probability) considered as a CE outcome. *AUC*, Area under the curve.

structure and small number of parameters are beneficial in the absence of a large training dataset, for example, those with SV physiology.

We analyzed the relationship between the time before CEs of patient states used to train machine-learning models and their performance at different prediction horizons, as shown in Figure 4. Models trained with data close to the time of CEs performed best at prediction of imminent deterioration, whereas models trained with data further from the time of CEs performed best in early prediction. For use in a clinical setting where having 1 single model may be a practical necessity, we found that the ensemble strategy produced a model with a reasonable balance between imminent and early prediction performance.

In recent years, there has been an increased interest in early warning systems and a variety of predictive models being used in clinical settings. These models are often validated in the adult population, and there have been efforts to adapt them to pediatric patients. However, the performance of generic early warning systems has suffered in certain populations in whom specific scores are better suited.[13,28,29] To the best of our knowledge, the model by Rusin and colleagues[21] is the only model available for real-time prediction of CEs in infants with SV. Although this model achieved higher prediction performance than our best model in the hour preceding CEs, clinicians may be aware of patient deterioration at that time. In comparison, our best model performed better at early prediction of CEs,

Rich Tsui, PhD

VIDEO 1. A brief overview of the C-WIN system, its innovation, and modeling strategies. Video available at: https://www.jtcvs.org/article/S0022-5223(19)30479-9/fulltext.

achieving moderate performance up to 8 hours before CEs. We achieved this using variables that are routinely available in most hospital EHR systems. Therefore, our models could be more feasibly implemented in institutions that do not have the technologic or financial resources required to collect and analyze data such as electrocardiogram waveforms in real time. Finally, we retrieved data from a larger cohort to develop our models, including 93 patients and 131 CEs.

Early, real-time prediction of CEs may help clinicians reduce morbidity, mortality, length of stay, health care costs, and the suffering of patients and guardians. However, to fully realize the benefits of implementing such systems in clinical practice, they should be specific enough to minimize alert fatigue and should not increase the burden of clinicians or divert resources from other aspects of care. From a clinical standpoint, our models may enable early interventions and avoidance of up to 49% of CEs with a specificity of 95% (based on performance 1 hour before CEs). This would allow clinicians to focus on patients truly at risk of CEs while minimizing alert fatigue. Furthermore, because our models use physiologic variables that may be extracted automatically from the EHR, an early warning system based on our models may operate autonomously and at low operational cost without disrupting clinicians' workflow.

We envision a potential implementation of our models as an alert-triggering system with a 2-tiered set of responses. First, a model calibrated for increased sensitivity and earliest response may be monitored via virtual surveillance in a Tele-ICU setting, following a well-established model in adult patients.[30] Then, a second model calibrated for high specificity may be used to prompt a rapid response from bedside clinicians.

Study Limitations

Our study had several limitations. First, all admissions came from a single institution (University of Pittsburgh Medical Center Children's Hospital of Pittsburgh). Although this center is one of the major tertiary referral

hospitals serving Western Pennsylvania and neighboring states (12,333 inpatient visits in 2017), our results may not be generalizable to different populations. Second, we could not retrospectively retrieve some variables that experts believed to be relevant, including electrocardiography, echocardiography, and x-ray imaging. However, those additional data may not be readily available in many hospitals and EHR systems. Nevertheless, it may be feasible to develop a more comprehensive model that leverages both routinely collected and high-frequency electrocardiogram data in the future. Thus, the same model may be applied in hospitals with varying levels of technologic resources. Third, we imputed missing values with data available within a 6-hour period. Although this may be a reasonable assumption, some variables may change rapidly, and more elaborate imputation techniques may improve performance. Fourth, our test set included a limited number of ECMO and CPR events, and we were not able to ascertain the performance of our models for these events separately. Fifth, our models did not leverage longitudinal changes in variables values, with the exception of changes in creatinine and mixed venous saturation. This will be an area for improvement in future iterations of our models. Likewise, we did not include variables related to clinical interventions. Although it is true that some interventions may signal imminent deterioration (eg, order of ADAMTS13 activity test), for the first iteration of our models we decided to focus exclusively on physiologic variables. Finally, event times and variable values collected from EHRs were often entered manually as part of routine care and may have been subject to potential data-entry errors.

CONCLUSIONS

We developed and validated predictive models that achieved early prediction of CEs in infants with SV physiology. These models used routinely collected data commonly available in hospital EHR systems. Thus, our models potentially can be deployed to most pediatric hospitals and help ICU teams to predict patient deterioration with up to 8 hours of anticipation. Early prediction of CEs may help clinicians perform interventions aimed at reducing morbidity, mortality, length of stay, and health care costs associated with such events.

CONG

References

1. Tabbutt S, Ghanayem N, Ravishankar C, Tabbutt S, Ghanayem N, Ravishankar C, Sleeper LA, Cooper DS, Frank DU, et al. Risk factors for hospital morbidity and mortality after the Norwood procedure: a report from the pediatric heart network single ventricle reconstruction trial. *J Thorac Cardiovasc Surg.* 2012;144:882-95.
2. Sluysmans T, Sanders SP, van der Velde M, Matitiau A, Parness IA, Spevak PJ, et al. Natural history and patterns of recovery of contractile function in single left ventricle after Fontan operation. *Circulation.* 1992;86:1753-61.
3. O'Leary PW. Prevalence, clinical presentation and natural history of patients with single ventricle. *Prog Pediatr Cardiol.* 2002;16:31-8.
4. Bouma BJ, Mulder BJM. Changing landscape of congenital heart disease. *Circ Res.* 2017;120:908-22.
5. Meza JM, Hickey EJ, Blackstone EH, Jaquiss RDB, Anderson BR, Williams WG, et al. The optimal timing of stage 2 palliation for hypoplastic left heart syndrome: an analysis of the pediatric heart network single ventricle reconstruction trial public data set. *Circulation.* 2017;136:1737-48.
6. Barron DJ, Kilby MD, Davies B, Wright JG, Jones TJ, Brawn WJ. Hypoplastic left heart syndrome. *Lancet.* 2009;374:551-64.
7. Kane-Gill SL, O'Connor MF, Rothschild JM, Selby NM, McLean B, Bonafide CP, et al. Technologic distractions (part 1): summary of approaches to manage alert quantity with intent to reduce alert fatigue and suggestions for alert fatigue metrics. *Crit Care Med.* 2017;45:1481-8.
8. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc.* 2006;13:138-47.
9. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. *Crit Care Med.* 1996;24:743-52.
10. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients*. *Crit Care Med.* 2006;34:1297-310.
11. Chapman SM, Wray J, Oulton K, Pagel C, Ray S, Peters MJ. "The Score Matters": wide variations in predictive performance of 18 paediatric track and trigger systems. *Arch Dis Child.* 2017;102:487-95.
12. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified early warning score in medical admissions. *QJM.* 2001;94:521-6.
13. da Silva YS, Hamilton MF, Horvat C, Fink EL, Palmer F, Nowalk AJ, et al. Evaluation of electronic medical record vital sign data versus a commercially available acuity score in predicting need for critical intervention at a tertiary children's hospital. *Pediatr Crit Care Med.* 2015;16:644-51.
14. Niles DE, Dewan M, Zebuhr C, Wolfe H, Bonafide CP, Sutton RM, et al. A pragmatic checklist to identify pediatric ICU patients at risk for cardiac arrest or code bell activation. *Resuscitation.* 2016;99:33-7.
15. Snoek KG, Capolupo I, Morini F, van Rosmalen J, Greenough A, van Heijst A, et al. Score for neonatal acute physiology-II predicts outcome in congenital diaphragmatic hernia patients. *Pediatr Crit Care Med.* 2016;17:540-6.
16. Moss TJ, Lake DE, Calland JF, Enfield KB, Delos JB, Fairchild KD, et al. Signatures of subacute potentially catastrophic illness in the ICU: model development and validation. *Crit Care Med.* 2016;44:1639-48.
17. Fenix JB, Gillespie CW, Levin A, Dean N. Comparison of pediatric early warning score to physician opinion for deteriorating patients. *Hosp Pediatr.* 2015;5:474-9.
18. Kennedy CE, Aoki N, Mariscalco M, Turley JP. Using time series analysis to predict cardiac arrest in a PICU. *Pediatr Crit Care Med.* 2015;16:e332-9.
19. Gupta P, Chakraborty A, Gossett JM, Rettiganti M. A prognostic tool to predict outcomes in children undergoing the Norwood operation. *J Thorac Cardiovasc Surg.* 2017;154:2030-7.e2.
20. Vu EL, Rusin CG, Penny DJ, Kibler KK, Easley RB, Smith B, et al. A novel electrocardiogram algorithm utilizing ST-segment instability for detection of cardiopulmonary arrest in single ventricle physiology: a retrospective study. *Pediatr Crit Care Med.* 2017;18:44-53.
21. Rusin CG, Acosta SI, Shekerdemian LS, Vu EL, Bavare AC, Myers RB, et al. Prediction of imminent, severe deterioration of children with parallel circulations using real-time processing of physiologic data. *J Thorac Cardiovasc Surg.* 2016;152:171-7.
22. Battiti R. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans Neural Networks.* 1994;5:537-50.
23. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *IJCAI'95 Proceedings of the 14th International Joint Conference on Artificial Intelligence.* Vol 2. San Francisco, CA: Morgan Kaufmann Publishers; 1995:1137-43.
24. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837-45.
25. Dunn OJ. Multiple comparisons among means. *J Am Stat Assoc.* 1961;56:52-64.
26. Warner HR, Toronto AF, Veasey LG, Stephenson R. A mathematical approach to medical diagnosis. *JAMA.* 1961;177:177.
27. López Pineda A, Ye Y, Visweswaran S, Cooper GF, Wagner MM, Tsui F. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *J Biomed Inform.* 2015;58:60-9.
28. Opio MO, Nansubuga G, Kellett J. Validation of the VitalPACTM early warning score (ViEWS) in acutely ill medical patients attending a resource-poor hospital in sub-Saharan Africa. *Resuscitation.* 2013;84:743-6.
29. Downey CL, Tahir W, Randell R, Brown JM, Jayne DG. Strengths and limitations of early warning scores: a systematic review and narrative synthesis. *Int J Nurs Stud.* 2017;76:106-19.
30. Lilly CM, McLaughlin JM, Zhao H, Baker SP, Cody S, Irwin RS, et al. A multicenter study of ICU telemedicine reengineering of adult critical care. *Chest.* 2014;145:500-7.

Key Words: cardiopulmonary resuscitation, congenital heart defects, endotracheal intubation, extracorporeal membrane oxygenation, hypoplastic left heart syndrome, risk assessment

CONG

TABLE E1. Diagnostic billing codes used to identify patients with single-ventricle physiology

Congenital heart disorder	Diagnostic code	Terminology
Common ventricle	745.3	ICD-9
Pulmonary valve atresia	746.01	ICD-9
	Q22.0	ICD-10
Tricuspid atresia and stenosis	746.1	ICD-9
Stenosis of aortic valve	746.3	ICD-9
Mitral stenosis	746.5	ICD-9
	Q23.2	ICD-10
Hypoplastic left heart syndrome	746.7	ICD-9
	Q22.6	ICD-10
Atresia and stenosis of aorta	747.22	ICD-9
Double inlet ventricle	Q20.4	ICD-10
Hypoplastic right heart syndrome	Q23.4	ICD-10

ICD-9, International Classification of Diseases, 9th Revision; *ICD-10*, International Classification of Diseases, 10th Revision.

TABLE E2. Model variables and value-selection priority levels

Model variable	EHR clinical concept	Priority
ADAMTS 13 activity	ADAMTS 13 activity	1
ALT	ALT/SGPT	1
Anti-Xa	Anti-Xa unfractionated heparin	1
	Anti-Xa assay for enoxaparin	1
Base excess*	Base excess, arterial	1
	Base excess, capillary	1
	Base excess, venous	1
	Base excess, oxygenator	1
	Base deficit oxygenator	1
	Base deficit capillary	1
	Base deficit, venous	1
	Base deficit arterial	1
	Base deficit venous	1
Bicarbonate anion*	HCO3a	1
	HCO3v	2
Blood urea nitrogen*	Blood urea nitrogen	1
Brain natriuretic peptide	B-type natriuretic peptide	1
Carbon dioxide	$Paco_2$	1
	$Pvco_2$	2
Central venous pressure*	Central venous pressure	1
Creatinine*	Cr	1
Creatinine change from baseline†	N/A	N/A
Diastolic BP*	Arterial diastolic pressure	1
	Diastolic BP	3
	Arterial diastolic pressure #2	2
Fibrinogen	Fibrinogen level	1
Fraction of inspired oxygen*	Fio_2	1
	Fio_2 (oxygen %)	2
Glucose	Glucose meter	1
	Glucose, whole blood	2
	Glucose	3
Heart rate*	heart rate	1
	heart rate—Spo_2	2
	Pulse	3
Hematocrit	Hct, whole blood	1
	Hct derived, venous	2
	Hct derived, arterial	2
	Hct	3
Hemoglobin	Hgb	1
	Hgb, venous	2
	Hgb, arterial	2
International normalized ratio	INR	1
Ionized calcium	Ionized Ca, whole blood	1
Lactate*	Lactate	1
	Lactate, whole blood	2
Mixed venous oxygen saturation*	O2 sat-mixed venous	1
	Svo_2	1
Mixed venous oxygen saturation change from baseline†	N/A	N/A

(Continued)

CONG

TABLE E2. Continued

Model variable	EHR clinical concept	Priority
Near-infrared spectroscopy*	NIRS cerebral Oxygenation-L	1
	NIRS cerebral oxygenation-R	1
	NIRS tissue oxygenation	2
	NIRS cerebral oxygenation	3
	NIRS cerebral oxygenation #2	3
Oxygen saturation*	S_{AO_2}	1
	Sp_{O_2} bedside monitor	1
Partial pressure of oxygen in arterial blood*	P_{AO_2}	1
Partial thromboplastin time	PTT	1
Platelets	Platelets	1
Potassium	K, whole blood	1
	K	2
Respiratory rate*	Respiratory rate	1
Sodium	Na, whole blood	1
	Na	2
Systolic BP*	Arterial systolic pressure	1
	Arterial systolic pressure #2	2
	Systolic BP	3
Total bilirubin	Bilirubin, total	1
	Bilirubin	1
Urine output mL/Hg/h*	Urine output 24-h (weight based)	1
	Urine output 8-h (weight based)	1

Model variables were identified by experienced cardiac-intensive-care-unit providers. The *EHR clinical concept* column shows the hospital-specific codes that represent variables in the dataset used for model development and evaluation. The *priority* values were used to select variable values when multiple EHR-event-code values were available simultaneously. *EHR*, Electronic health record; *ALT*, alanine transaminase; *SGPT*, serum glutamic-pyruvic transaminase; *N/A*, not applicable; *BP*, blood pressure; *F_{IO_2}*, inspired oxygen fraction. *Variables included in the minimal subset that experts identified as essential for the prediction of CEs. †Variable derived from another variable in this table.