Published in final edited form as:

Pediatr Crit Care Med. 2017 May ; 18(5): 469–476. doi:10.1097/PCC.0000000000001150.

Integration of Single-Center Data-Driven Vital Sign Parameters into a Modified Pediatric Early Warning System

Catherine E. Ross, M.D.¹, Iliana J. Harrysson, M.D.², Veena V. Goel, M.D.³, Erika J. Strandberg, M.S.⁴, Peiyi Kan, M.S.⁵, Deborah E. Franzon, M.D.⁶, and Natalie M. Pageler, M.D., M.Ed.^{7,8,9}

¹Division of Medicine Critical Care, Department of Medicine, Boston Children's Hospital, Boston, MA, USA

²Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

³Department of Pediatrics, Sutter Health Palo Alto Medical Foundation, Palo Alto, CA, USA

⁴Biomedical Informatics, Stanford University School of Medicine, Stanford, CA, USA

⁵Statistical Unit, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

⁶Division of Critical Care Medicine, Department of Pediatrics, University of California, San Francisco, San Francisco, CA, USA

⁷Division of Critical Care Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

⁸Department of Clinical Informatics, Stanford Children's Health, Stanford, CA, USA

⁹Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

Abstract

Objective—Pediatric early warning systems using expert-derived vital sign parameters demonstrate limited sensitivity and specificity in identifying deterioration. We hypothesized that modified tools using data-driven vital sign parameters would improve the performance of a validated tool.

Design—Retrospective Case Control

Setting—Quaternary-care children's hospital

Patients—Hospitalized, non-critically ill patients less than 18 years old. Cases were defined as patients who experienced an emergent transfer to an intensive care unit (ICU) or out-of-ICU

Author Correspondence to: Catherine Ross, 333 Longwood Ave, Division of Medicine Critical Care, Boston, MA, 02115, Catherine.Ross@childrens.harvard.edu.

Reprints: Reprints will not be ordered.

Institution: This study was performed at Lucile Packard Children's Hospital at Stanford.

Copyright form disclosure: The remaining authors have disclosed that they do not have any potential conflicts of interest.

cardiac arrest. Controls were patients who never required intensive care. Cases and controls were split into training and testing groups.

Intervention—The Bedside Pediatric Early Warning System (bPEWS) was modified by integrating data-driven heart rate and respiratory rate parameters (modified bPEWS 1 and 2). Modified bPEWS 1 used the 10th and 90th percentiles as normal parameters while modified bPEWS 2 used 5th and 95th percentiles.

Measurements and Main Results—The training set consisted of 358 case events and 1830 controls; the testing set had 331 case events and 1215 controls. In the sensitivity analysis, 207 of the 331 testing set cases (62.5%) were predicted by the original tool versus 206 (62.2%, $p=0.54$) with modified bPEWS 1 and 191 (57.7%, $p<0.001$) with modified bPEWS 2. For specificity, 1005 of the 1215 testing set control patients (82.7%) were identified by original bPEWS versus 1013 (83.1%, $p=0.54$) with modified bPEWS 1 and 1055 (86.8%, $p<0.001$) modified bPEWS 2. There was no net gain in sensitivity and specificity using either of the modified bPEWS tools.

Conclusions—Integration of data-driven vital sign parameters into a validated pediatric early warning system did not significantly impact sensitivity or specificity, and all the tools showed lower than desired sensitivity and specificity at a single cutoff point. Future work is needed to develop an objective tool that can more accurately predict pediatric decompensation.

Keywords

Pediatrics; critical illness; cardiac arrest; respiratory rate; heart rate; hospital medicine

INTRODUCTION

Several pediatric early warning tools have been designed to predict in-hospital deterioration (1–12) based on literature showing that a large percentage of adult and pediatric patients who suffer in-hospital cardiopulmonary arrest (CPA) have detectable signs of clinical deterioration in the 24–48 hours prior to arrest (13–20). Few of these tools have been adequately validated (21), and even fewer have demonstrated sufficient sensitivity and specificity in predicting deterioration to meaningfully guide early intervention (21, 22). This highlights the need for development and validation of a more accurate early warning tool.

One identified area of weakness in published pediatric early warning tools is the lack of objectivity in age-specific vital sign parameters, which are universally either selected by expert opinion (21, 23) or not delineated at all. Fleming et al. were the first to demonstrate that measured vital signs in well children differed from traditional normal ranges found in reference texts (24). Bonafide et al. went on to show that measured heart rate (HR) and respiratory rate (RR) values in hospitalized, non-critically ill children differ from traditional pediatric vital sign reference ranges, and many of these values would fall outside of the normal ranges used in published early warning tools (23). We hypothesized that a novel approach of integrating data-driven HR and RR parameters into a validated pediatric early warning system would improve the tool's sensitivity and specificity in predicting emergent intensive care unit (ICU) transfers and out-of-ICU codes.

MATERIALS AND METHODS

We performed this retrospective case-control study on the general medical and surgical units at Lucile Packard Children's Hospital Stanford (LPCHS). LPCHS is a 302-bed quaternary-care academic children's hospital that offers a full complement of medical and surgical subspecialties and robust transplant programs. During the study period, LPCHS used the Cerner electronic medical record (EMR) system (Millennium, Cerner, Kansas City, MO, USA). This study was exempted by the Stanford University Institutional Review Board.

Development of the Modified bPEWS 1 and 2

At our institution, Goel et al. (25) collected HR and RR measurements on all non-critically ill hospitalized patients from January 1st, 2013 to May 3rd, 2014 and developed percentile curves with methods similar to those used by Bonafide et al. (23). Defined cutoffs at the 1st, 2.5th, 5th, 10th, 90th, 95th, 97.5th and 99th percentiles for HR and RR in each age group were calculated. Based on these percentiles, the Bedside Pediatric Early Warning System (bPEWS) (8, 26) was modified to include new cutoff points for HR and RR. Because the optimal percentile cutoffs for this purpose were unknown, we developed two modified bPEWS systems using different percentiles to allocate scores for HR and RR, with increasing points as they deviate further from the mean. For the first system, "modified bPEWS 1", we defined the 10th/90th, 5th/95th and 1st/99th percentiles as cutoffs for 1, 2 and 4 points, respectively. "Modified bPEWS 2" was defined as the 5th/95th, 2.5th/97.5th and 1st/99th percentiles for the same point allocation. Additionally, the 5 age groups in bPEWS were further subdivided into 13 age groups given the observation that the change in measured HR and RR with age is greatest from zero to five years (23, 24). The bPEWS tool was otherwise preserved and scored according to the published tool.

Respiratory Effort Scoring

Because historical nursing documentation of respiratory effort evaluation consisted of qualitative descriptions rather than the numerical rating in bPEWS, we collected additional data subsequent to the study period to develop a predictive model that estimated a numerical value on the bPEWS respiratory effort scale. A detailed description of the development of the predictive model (Supplemental Digital Content Text 1) and score distribution (Supplemental Digital Content Table 1) can be found in the supplementary materials. The final proportional odds model was used to predict respiratory effort scores for the entire data set. Analysis was performed in R (27) using the MASS (28) and Soft-Impute (29) packages. Additionally, as per the original bPEWS model, if any patient had the word "Apnea" they were assigned a respiratory effort score of 4.

Patient Selection

The inclusion criteria for both cases and controls were patients <18 years of age admitted between January 1st, 2008 and May 5th, 2014 who had at least one set of documented vital signs on the pediatric general care floor. Exclusion criteria consisted of patients with limitation of care orders in place during the period studied and patients who spent no time on the general care floor. Vital signs and assessments documented while in the pediatric ICU, pediatric cardiovascular ICU, neonatal ICU, outpatient areas, emergency department or

procedural/surgical areas during the study period were not included. A patient may have spent time in these areas during other parts of the hospitalization, but only vital signs documented while on the general care floor were used to calculate early warning scores. There were no exclusions based on gender or race.

Case events were selected by manual chart review of all patients who were transferred to a critical care unit from any general medical/surgical floor. All charts were reviewed by two independent reviewers on the investigating team and all discrepancies were resolved with a third review. The event was determined to be a case if the transfer was unplanned, emergent and the patient received an ICU-level intervention within 24 hours of transfer to ICU. ICU-level interventions were defined by our hospital policies delineating therapies performed exclusively in the ICU or requiring immediate ICU transfer upon patient stabilization as follows: cardiopulmonary resuscitation; bag-mask ventilation; emergent intubation; status epilepticus requiring two or more anti-epileptic drug loads or with respiratory compromise; use of vasoactive infusions with the exception of our oncology patients who are able to receive dopamine up to 8mcg/kg/min on the general care floor and heart failure patients who can receive milrinone infusion after being initiated in the ICU; respiratory support of high flow nasal cannula >4L/min or any positive pressure ventilation greater than the patient's baseline; status asthmaticus requiring continuous albuterol >6 hours; anti-arrhythmic infusions or cardioversion; unexpected, emergent surgical interventions requiring post-operative ICU care; stroke, posterior reversible encephalopathy syndrome, measured intracranial hypertension and other symptomatic neurologic decompensation with objective evidence by imaging. These criteria were selected to ensure that case events represented true clinical deterioration in the general care setting and intentionally excluded transfers solely for intensive nursing care, closer observation, routine post-operative care or sedation/procedural needs. Transfers that did not meet one or more of these criteria were not used as case events.

Controls were defined as all other patients admitted to the general care floor who did not spend any time in an ICU during the admission. In consideration of logistical constraints, controls were randomly selected in a 4:1 ratio according to standard statistical methodology (30). Additionally, the controls were selected in an unmatched fashion so as to represent the entire at-risk population of hospitalized children (31).

Data Collection and Scoring

To target the time period leading up to decompensation, the case events were studied for a 23-hour period beginning 24 hours and ending one hour prior to their emergent ICU transfer or code. As suggested by Chapman et al. (21), data collected in the hour immediately prior to the event was not used in the analysis as a critical score in this time period would not provide meaningful lead time for an intervention. Controls were studied for the entirety of their admission to ensure that the peak of illness was captured. All recorded data for the seven categories in bPEWS were extracted from the LCHS clinical data warehouse, and the bPEWS and modified bPEWS 1 and 2 scores were calculated for each discrete time of documentation. If a value for one or more categories was missing for a given time, the last value for that category was carried forward for up to 12 hours. If no value was ever recorded

for a category, it was presumed to be within normal limits and assigned a score of zero. With the exception of HR and RR, scores for the five remaining categories for a given patient at a given time were the same in all three warning systems. The highest cumulative score for each case/control was used for the statistical analysis.

Statistical Analysis

Case and control patients admitted between 2008 and 2011 were assigned to the training set, and those admitted between 2012 and 2014 were assigned to the testing set (Supplemental Digital Content Figure 1). The training set was used to calculate the area under the receiver operating characteristic (AUROC) curve and the optimal cutoff point for each tool, which was defined by Youden's J point (the point on the ROC curve which is farthest from line of random chance). The remainder of the analysis was performed on the testing set. One-way ANOVA was used to compare the average maximum scores between the case and control group for each tool. Sensitivity and specificity for each tool were calculated using their respective optimal cutoff scores, and were compared between the tools with McNemar's test. Cohen's Kappa was calculated to determine the degree of agreement between the original bPEWS and the modified systems in identifying cases and controls. The mean time prior to transgression of critical score was calculated for the case events.

Sub-group analyses were performed based on age less than four years because one of the major modifications in modified bPEWS 1 and 2 was to further sub-divide the three age groups for under four years in the original bPEWS into eight age groups. Further sub-group analysis was performed on cases who experienced an out-of-ICU code (CPA or respiratory arrest), circulatory shock or respiratory insufficiency. These groups were chosen based on the hypothesis that HR and RR derangements may be more prominent in these disease processes and therefore more likely to benefit from data-driven vital sign parameters.

Finally, in a post-hoc analysis, we used the HR and RR parameters in modified bPEWS 1 and 2 in isolation to determine if the data-driven parameters in either category alone would improve the tool as compared to combined HR and RR modifications. We calculated optimal cutoffs, AUROC, sensitivity and specificity using identical methods as detailed above.

RESULTS

There were 689 emergent ICU transfers including 83 out-of-ICU codes that met inclusion criteria. The training set consisted of 358 case events and 1830 controls; the testing set had 331 case events and 1215 controls. Baseline characteristics are shown in Table 1. The AUROC curve were 0.818, 0.812 and 0.809 for original bPEWS, modified bPEWS 1 and modified bPEWS 2, respectively, with optimal cutoff scores of 7, 8 and 8, respectively (Figure 1).

There was a significant difference in the mean maximum scores for cases compared to controls in all three tools (Table 2). In the sensitivity analysis, 207 of the 331 testing set cases (62.5%) would be predicted by the original tool versus 206 (62.2%, $p=0.54$) with modified bPEWS 1 and 191 (57.7%, $p<0.001$) for modified bPEWS 2. For specificity, 1005 of the 1215 testing set control patients (82.7%) were correctly identified by original bPEWS

versus 1013 (83.1%, $p=0.54$) with modified bPEWS 1 and 1055 (86.8%, $p<0.001$) modified bPEWS 2 (Table 3). The time to transgression of the critical cutoff was similar for original bPEWS, modified bPEWS 1 and modified bPEWS 2 at 10.7 hours, 10.9 hours and 10.2 hours, respectively.

In the sub group analysis, there was no net gain in sensitivity and specificity in children less than four years old with either of the modified tools (Table 4). Additionally, modified bPEWS 1 had slightly better sensitivity in detecting patients who went on to have respiratory insufficiency. Otherwise, both novel tools were equivalent or inferior to original bPEWS in detecting patients who had out-of-ICU codes and circulatory shock (Table 4).

The results of the post-hoc analysis showed no significant net increase in sensitivity and specificity using data-driven RR parameters in isolation. Using data-driven HR parameters in isolation slightly decreased the efficacy of the tool.

DISCUSSION

To our knowledge, this is the first pediatric early warning system to use data-driven HR and RR parameters. In our retrospective case control study, integration of data-driven vital sign parameters into a validated tool did not significantly impact the sensitivity, specificity or lead-time in predicting pediatric in-hospital decompensation. Furthermore, at a cutoff score of seven, our study showed lower specificity for the original bPEWS compared to results of a multi-center validation study (83% vs. 91%, respectively) despite similar sensitivity (64% vs. 63%, respectively) (26). This is likely due to differences in sampling methods of the control populations given Parshuram et al. examined only 12 hours of data for each control patient compared to the current study, which spans the entire admission. Our methods capture the peak severity of illness for every control patient which more accurately conveys the burden of false positives if implemented into a rapid response triggering system.

Additionally, we expected that younger children would benefit more from the modified bPEWS tools given that the vital sign parameters were more specific to age with the addition of more age group categories under age four. We also theorized that respiratory insufficiency, circulatory shock and out-of-ICU codes would have more HR and RR derangements and therefore would have improved sensitivity with our modifications. However, sub-group analysis did not show significant improvement for any of these groups.

Our findings highlight the difficulty in the objective assessment of pediatric deterioration. In a systematic review, Chapmen et al. (21) found only three validated pediatric tools (3, 12, 32) that accurately reported sensitivity and specificity, none of which exceeded 85% for both at a single cutoff score. Additionally, sensitivity was likely over-estimated in two of these studies (12, 32) owing to methods which included studying the time of the serious adverse event which do not provide significant lead time for early intervention. Since then, several novel or modified tools have been developed and tested (9, 26, 33, 34) which also failed to demonstrate sensitivity and specificity greater than 85% at a single cutoff score.

In 2005, Monaghan published the first pediatric early warning score, “Brighton PEWS” (7), at a time when charting was still largely done by hand. For this reason, delineation of

specific age-based vital sign parameters was intentionally omitted for ease and accuracy of real-time bedside scoring. In line with this thinking, a large portion of published pediatric early warning tools reference a degree of deviation from “normal”, but do not define normal limits (6, 7, 9, 12, 35). This introduces an element of subjectivity for nearly all the sub-categories of these tools, therefore limiting their generalizability to outside institutions. However, in the age of the EMR, the need for simplicity is less limiting. An automated tool could provide accurate scoring which does not require additional nursing documentation time and can be updated in real-time with every new piece of data that is entered. These more complex algorithms have been shown to be effective in predicting outcomes in adults (36). However pediatric studies thus far, although promising, lack similar success (33, 37).

The accuracy of a purely objective early warning system is inherently limited by the overlap in distribution of physiologic and pathologic vital sign changes. This holds true for both inter- and intra-patient variability. Even when these parameters are tailored to a population, as in the current study, there is no improvement in detecting deterioration. Therefore, future work should examine vital sign trends and deviation from a patient’s own baseline in an individualized approach as opposed to vital sign parameters derived from the population. Alternatively, data-driven vital sign parameters may be useful if tailored to a more specific population (i.e. by disease process).

Our major limitation was missing data in relation to the numerical scoring of respiratory effort. This was mitigated by a predictive model which likely underestimated the score in this category (Supplemental Digital Content Table 1). However, the predicted respiratory effort category sub-score for every patient was the same in all three early warning systems and therefore did not bias the study toward one tool over the others. Additionally, we were unable to capture control patients who spent time in the ICU and were later transferred to the general care floor but did not experience a deterioration. For this reason, patients admitted for cardiac surgery or solid organ transplant who received routine post-operative care in the ICU were likely underrepresented in our control population. Given the high variability in physiologically normal vital signs in the post-operative cardiac population, this may have led to an overestimation in specificity for all the tools in this group. Another limitation in our study and PEWS literature in general, is the assumption that all in-hospital critical events are preceded by measurable physiologic changes. However, many events such as seizure or aspiration may truly be unpredictable and may not be appropriate to use as case patients. Finally, this study was performed at a single institution and may not be generalizable to institutions with different patient populations.

CONCLUSION

In summary, incorporating population-based data-driven HR and RR parameters into a validated pediatric early warning system had no effect on the performance of bPEWS. Neither the original tool nor the modified tools had high sensitivity and specificity at a single cutoff point to delineate sick from well hospitalized children. Future work should focus on a more sophisticated tool design integrating individualized vital sign trends to provide more granularity without over-burdening frontline care providers with false alarms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Debby Huang, M.S., Lynda Knight R.N., CCRN and the LPCHS code committee for their assistance in data retrieval; Suzan Carmichael, PhD for her assistance in statistical support; and Tua Palangyo, R.N. and Amy Chapman, R.N. for their assistance in respiratory effort nursing surveys.

Funding Source: This project was funded by Jackson Vaughan Critical Care Research Fund; Ms. Strandberg's contribution was funded by the National Library of Medicine Informatics Training Grant funded by the NIH

Dr. Ross' institution received funding from Jackson Vaughan Critical Care Research Fund and from National Library of Medicine Informatics Training Grant funded by the NIH. She received support for article research from the NIH, and disclosed: Ms. Strandberg's work is funded by National Library of Medicine Informatics Training Grant funded by the NIH. Ms. Strandberg received support for article research from the NIH. Her institution received funding from National Library of Medicine Informatics Training Grant. Dr. Pageler's institution received funding from Jackson Vaughan Critical Care Research Fund.

References

1. Bell D, Mac A, Ochoa Y, et al. The Texas Children's Hospital Pediatric Advanced Warning Score as a predictor of clinical deterioration in hospitalized infants and children: a modification of the PEWS tool. *J Pediatr Nurs.* 2013; 28(6):e2–9. [PubMed: 23685263]
2. Bonafide CP, Holmes JH, Nadkarni VM, et al. Development of a score to predict clinical deterioration in hospitalized children. *J Hosp Med.* 2012; 7(4):345–349. [PubMed: 22489072]
3. Duncan H, Hutchison J, Parshuram CS. The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalized children. *J Crit Care.* 2006; 21(3):271–278. [PubMed: 16990097]
4. Egdel P, Finlay L, Pedley DK. The PAWS score: validation of an early warning scoring system for the initial assessment of children in the emergency department. *Emerg Med J.* 2008; 25(11):745–749. [PubMed: 18955610]
5. Haines C, Perrott M, Weir P. Promoting care for acutely ill children-development and evaluation of a paediatric early warning tool. *Intensive Crit Care Nurs.* 2006; 22(2):73–81. [PubMed: 16271295]
6. McLellan MC, Connor JA. The Cardiac Children's Hospital Early Warning Score (C-CHEWS). *J Pediatr Nurs.* 2013; 28(2):171–178. [PubMed: 22903065]
7. Monaghan A. Detecting and managing deterioration in children. *Paediatr Nurs.* 2005; 17(1):32–35. [PubMed: 15751446]
8. Parshuram CS, Hutchison J, Middaugh K. Development and initial validation of the Bedside Paediatric Early Warning System score. *Crit Care.* 2009; 13(4):R135. [PubMed: 19678924]
9. Skaletzky SM, Raszyński A, Totapally BR. Validation of a modified pediatric early warning system score: a retrospective case-control study. *Clin Pediatr (Phila).* 2012; 51(5):431–435. [PubMed: 22157421]
10. Solevag AL, Eggen EH, Schroder J, et al. Use of a modified pediatric early warning score in a department of pediatric and adolescent medicine. *PLoS One.* 2013; 8(8):e72534. [PubMed: 23991121]
11. Tibballs J, Kinney S, Duke T, et al. Reduction of paediatric in-patient cardiac arrest and death with a medical emergency team: preliminary results. *Arch Dis Child.* 2005; 90(11):1148–1152. [PubMed: 16243869]
12. Tucker KM, Brewer TL, Baker RB, et al. Prospective evaluation of a pediatric inpatient early warning scoring system. *J Spec Pediatr Nurs.* 2009; 14(2):79–85. [PubMed: 19356201]
13. Buist M, Bernard S, Nguyen TV, et al. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation.* 2004; 62(2):137–141. [PubMed: 15294398]

14. Buist MD, Jarmolowski E, Burton PR, et al. Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. *Med J Aust.* 1999; 171(1):22–25. [PubMed: 10451667]
15. Goldhill DR, White SA, Sumner A. Physiological values and procedures in the 24 h before ICU admission from the ward. *Anaesthesia.* 1999; 54(6):529–534. [PubMed: 10403864]
16. McQuillan P, Pilkington S, Allan A, et al. Confidential inquiry into quality of care before admission to intensive care. *BMJ.* 1998; 316(7148):1853–1858. [PubMed: 9632403]
17. Sax FL, Charlson ME. Medical patients at high risk for catastrophic deterioration. *Crit Care Med.* 1987; 15(5):510–515. [PubMed: 3568715]
18. Schein RM, Hazday N, Pena M, et al. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest.* 1990; 98(6):1388–1392. [PubMed: 2245680]
19. Smith AF, Wood J. Can some in-hospital cardio-respiratory arrests be prevented? A prospective survey. *Resuscitation.* 1998; 37(3):133–137. [PubMed: 9715771]
20. Tume L. The deterioration of children in ward areas in a specialist children's hospital. *Nurs Crit Care.* 2007; 12(1):12–19. [PubMed: 17883659]
21. Chapman SM, Grocott MP, Franck LS. Systematic review of paediatric alert criteria for identifying hospitalised children at risk of critical deterioration. *Intensive Care Med.* 2010; 36(4):600–611. [PubMed: 19940976]
22. Robson MA, Cooper CL, Medicus LA, et al. Comparison of three acute care pediatric early warning scoring tools. *J Pediatr Nurs.* 2013; 28(6):e33–41. [PubMed: 23276507]
23. Bonafide CP, Brady PW, Keren R, et al. Development of heart and respiratory rate percentile curves for hospitalized children. *Pediatrics.* 2013; 131(4):e1150–1157. [PubMed: 23478871]
24. Fleming S, Thompson M, Stevens R, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *Lancet.* 2011; 377(9770):1011–1018. [PubMed: 21411136]
25. Goel VV, Poole SF, Longhurst CA, et al. Safety analysis of proposed data-driven physiologic alarm parameters for hospitalized children. *J Hosp Med.* 2016; 11(12):817–823. [PubMed: 27411896]
26. 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(4):R184. [PubMed: 21812993]
27. R Core Development Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.
28. Venables, WN., Ripley, BD. *Modern Applied Statistics with S.* Fourth. New York: Springer; 2002.
29. Hastie T, Masumder R. Soft-Impute: matrix completion via iterative soft-thresholding svd. 2013
30. Taylor JM. Choosing the number of controls in a matched case-control study, some sample size, power and efficiency considerations. *Stat Med.* 1986; 5(1):29–36. [PubMed: 3961313]
31. Rose S, Laan MJ. Why match? Investigating matched case-control study designs with causal effect estimation. *Int J Biostat.* 2009; 5(1) Article 1.
32. Edwards ED, Powell CV, Mason BW, et al. Prospective cohort study to test the predictability of the Cardiff and Vale paediatric early warning system. *Arch Dis Child.* 2009; 94(8):602–606. [PubMed: 18812403]
33. da Silva YS, Fiedor Hamilton M, Horvat C, 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
34. McLellan MC, Gauvreau K, Connor JA. Validation of the Cardiac Children's Hospital Early Warning Score: An Early Warning Scoring Tool to Prevent Cardiopulmonary Arrests in Children with Heart Disease. *Congenit Heart Dis.* 2013
35. Akre M, Finkelstein M, Erickson M, et al. Sensitivity of the pediatric early warning score to identify patient deterioration. *Pediatrics.* 2010; 125(4):e763–769. [PubMed: 20308222]
36. Rothman MJ, Rothman SI, Beals Jt. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. *J Biomed Inform.* 2013; 46(5):837–848. [PubMed: 23831554]

37. Zhai H, Brady P, Li Q, et al. Developing and evaluating a machine learning based algorithm to predict the need of pediatric intensive care unit transfer for newly hospitalized children. Resuscitation. 2014

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Comparison of receiver operating characteristic curves.

Table 1

Patient characteristics.

Characteristic	Cases (n=689)		Controls (n=3045)	
	Training (n=358)	Testing (n=331)	Training (n=1830)	Testing (n=1215)
Mean Age at Admission in Years (SEM)	6 (0.30)	5.6 (0.29)	7.5 (0.13)	7.7 (0.17)
Age Group at Admission, No. (%)				
<3mo	27 (7.5)	26 (7.9)	74 (4.0)	66 (5.4)
3-<6mo	28 (7.8)	19 (5.7)	66 (3.6)	51 (4.2)
6-<9mo	15 (4.2)	19 (5.7)	47 (2.6)	34 (2.8)
9-<12mo	15 (4.2)	11 (3.3)	58 (3.2)	34 (2.8)
12-<18mo	32 (8.9)	25 (7.6)	97 (5.3)	72 (5.9)
18-<24mo	17 (4.7)	12 (3.6)	92 (5.0)	45 (3.7)
2-<3yr	31 (8.7)	35 (10.6)	150 (8.2)	82 (6.7)
3-<4yr	18 (5.0)	29 (8.8)	110 (6.0)	59 (4.9)
4-<6yr	26 (7.3)	34 (10.3)	211 (11.5)	117 (9.6)
6-<8yr	37 (10.3)	24 (7.3)	162 (8.9)	105 (8.6)
8-<12yr	35 (9.8)	42 (12.7)	237 (13.0)	186 (15.3)
12-<15yr	30 (8.4)	29 (8.8)	232 (12.7)	170 (14.0)
15-<18yr	47 (13.1)	26 (7.9)	294 (16.1)	194 (16.0)
Gender, No. (%)				
Female	152 (42.5)	133 (40.2)	872 (47.7)	584 (48.1)
Male	206 (57.5)	198 (59.8)	958 (52.3)	631 (51.9)
Primary Service on Admission, No. (%)				
Cardiology	80 (22.3)	61 (18.4)	91 (5.0)	67 (5.5)
Gastroenterology	22 (6.1)	16 (4.8)	133 (7.3)	86 (7.1)
General Pediatrics/Neurology	56 (15.6)	64 (19.3)	538 (29.4)	401 (33.0)
Hematology/Oncology/BMT	67 (18.7)	51 (15.4)	398 (21.7)	252 (20.7)
Nephrology	9 (2.5)	13 (3.9)	58 (3.2)	56 (4.6)
Pulmonology	47 (13.1)	43 (13.0)	58 (3.2)	52 (4.3)
Other Medical	66 (18.4)	75 (22.7)	34 (1.9)	25 (2)
Surgical	11 (3.1)	8 (2.4)	520 (28.4)	276 (22.7)
Intervention Required for Cases, No. (%)				
Respiratory Support	220 (61.5)	206 (62.2)	n/a	n/a
Vasopressors	50 (14.0)	60 (18.1)	n/a	n/a
Other Vasoactives	9 (2.5)	6 (1.8)	n/a	n/a
Transfusion for Acute Hemorrhage	13 (3.6)	10 (3.0)	n/a	n/a

Characteristic	Cases (n=689)		Controls (n=3045)	
	Training (n=358)	Testing (n=331)	Training (n=1830)	Testing (n=1215)
Anti-Arrhythmic Infusion / Cardioversion	17 (4.7)	12 (3.6)	n/a	n/a
Multiple Anti-epileptics or Airway Management for Status Epilepticus	21 (5.9)	14 (4.2)	n/a	n/a
Interventions for Other Neurologic Emergencies	19 (5.3)	18 (5.4)	n/a	n/a
Out-of-ICU Bag Mask Ventilation or CPR	52 (14.5)	31 (9.4)	n/a	n/a
CPR or Emergent Intubation <24 Hours After Transfer	94 (26.3)	49 (14.8)	n/a	n/a
Other	9 (2.5)	5 (1.5)	n/a	n/a
Number of Total Scores Generated (Mean Per Case/Control)	7077 (19.8)	5899 (17.8)	112321 (61.4)	72749 (59.9)
Number of Sub-Category Observations (Mean Per Case/Control)				
Heart Rate	5688(15.9)	4543 (13.7)	86454 (47.2)	51716 (42.6)
Respiratory Rate	5403 (15.1)	4533 (13.7)	61360 (33.5)	41074 (33.8)
Systolic Blood Pressures	3306 (9.2)	2613 (7.9)	69692 (38.1)	48870 (40.2)
Oxygen Saturation	6049 (16.9)	4933 (14.9)	82457 (45.1)	47323 (38.9)
Respiratory Effort	1767 (4.9)	1716 (5.2)	35084 (19.2)	23354 (19.2)
Oxygen Therapy	3491 (9.8)	2772 (8.4)	53504 (29.2)	31817 (26.2)
Capillary Refill Time	1435 (4.0)	1272 (3.8)	35163 (19.2)	22091 (18.2)

Abbreviations: SEM, Standard Error Measure; Mo, months; Yr, years; BMT, Bone Marrow Transplant; ICU, intensive care unit; CPR, cardio-pulmonary resuscitation

Table 2

Comparison of mean maximum scores of cases and controls in each system.

System	Mean Maximum Score (SEM)		
	Cases (n=331)	Controls (n=1215)	p-Value
Original bPEWS	7.7 (0.17)	4.5 (0.06)	<0.001
Modified bPEWS 1	8.5 (0.19)	4.9 (0.07)	<0.001
Modified bPEWS 2	8.1 (0.20)	4.4 (0.08)	<0.001

Abbreviations: SEM, Standard Error Measure; bPEWS, Bedside Pediatric Early Warning System

Table 3

Comparison of sensitivity, specificity and AUROC.

System	Optimal Cutoff Score	Sensitivity	Specificity	AUROC Curve	Cohen's Kappa (95% CI)	p-Value
Original bPEWS	7	0.63	0.83	0.818	Reference	Reference
Modified bPEWS 1	8	0.62	0.83	0.812	0.71 (0.67–0.75)	0.54
Modified bPEWS 2	8	0.58	0.87	0.809	0.70 (0.66–0.74)	<0.001

Abbreviations: AUROC, area under receiver operating characteristic; CI, Confidence Interval; bPEWS; Bedside Pediatric Early Warning System

Table 4

Comparison of sensitivity and specificity in sub-group analysis.

Subgroup	Original bPEWS		Modified bPEWS 1		Modified bPEWS 2	
	Sensitivity No. (%)	Specificity No. (%)	Sensitivity No. (%)	Specificity No. (%)	Sensitivity No. (%)	Specificity No. (%)
Age <4yr (n=172 Cases, 443 Controls)	109 (63)	357 (81)	112 (65)	341 (77)	102 (59)	368 (83)
Respiratory Insufficiency (n=206 Cases)	153 (74.3)	n/a	160 (77.8)	n/a	147 (71.4)	n/a
Circulatory Shock (n=60 Cases)	34 (57)	n/a	27 (45)	n/a	27 (45)	n/a
Out-of-ICU Code (n=31 Cases)	18 (58)	n/a	18 (58)	n/a	13 (42)	n/a

Abbreviations: bPEWS, Bedside Pediatric Early Warning System