Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

External Validation of the Ped-CMU Extubation Failure Predictive Score and Clinician Adherence in Pediatric Cardiac Patients

Thamonwan Kaewloon
Division of Cardiology, Department of Pediatrics, Faculty of Medicine, Chiang Mai University

Nop Khongthon
Center for Clinical Epidemiology and Clinical Statistics, Department of Biomedical Informatics and Clinical Epidemiology (BioCE), Faculty of Medicine

Pakpoom Wongyikul
Center for Clinical Epidemiology and Clinical Statistics, Department of Biomedical Informatics and Clinical Epidemiology (BioCE), Faculty of Medicine

Phichayut Phinyo
Center for Clinical Epidemiology and Clinical Statistics, Department of Biomedical Informatics and Clinical Epidemiology (BioCE), Faculty of Medicine

Kwannapas Saengsin
Kwannapas_09@hotmail.com

Division of Cardiology, Department of Pediatrics, Faculty of Medicine, Chiang Mai University

Article

Keywords: Extubation failure, Clinical prediction model, External validation, Paediatric cardiac surgery, Paediatric intensive care

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10051211/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

External Validation of the Ped-CMU Extubation Failure Predictive Score and Clinician Adherence in Pediatric Cardiac Patients

Thamonwan Kaewloon[1,‡]; Nop Khongthon, M.D.[2,‡]; Pakpoom Wongyikul, M.D.[2]; Phichayut Phinyo, M.D., Ph.D.[2]; Kwannapas Saengsin, M.D., Ph.D.[1,*]

- [1] Division of Cardiology, Department of Pediatrics, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand.
- [2] Center for Clinical Epidemiology and Clinical Statistics, Department of Biomedical Informatics and Clinical Epidemiology (BioCE), Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand.

‡ These authors contributed equally to this work.

Corresponding author:

Kwannapas Saengsin, M.D. Ph.D.

Division of Cardiology, Department of Pediatrics

Faculty of Medicine, Chiang Mai University

Chiang Mai, Thailand.

E-mail: Kwannapas_09@hotmail.com

Telephone: (66) 957813706

Abstract

Extubation failure remains a significant challenge in pediatric cardiac intensive care and is associated with increased morbidity and prolonged hospitalization. The Pediatric Chiang Mai University Extubation Failure Predictive Score (Ped-CMU ExFPS) was developed to identify cardiac patients at increased risk, but requires external validation before clinical implementation. We externally validated the Ped-CMU ExFPS in a contemporary cohort. In this retrospective study, pediatric patients aged 1 month to 18 years with congenital or acquired heart disease who underwent extubation in a Pediatric Cardiac Intensive Care Unit between June 2022 and June 2024 were included, with each extubation attempt analyzed as an independent episode. Model discrimination, calibration, and diagnostic indices were assessed. A total of 142 validation episodes were compared with 352 development episodes. The incidence of extubation failure was lower in the validation cohort (7.7% versus 11.0%). The score showed acceptable discrimination (AUROC 0.77, 95% CI 0.61-0.93), comparable to development (0.76), with reasonable calibration. A history of re-intubation remained a strong predictor (odds ratio 12.0, 95% CI 2.53-67.2). Sensitivity (81.8%) and negative predictive value (97.8%) were high, whereas positive predictive value was modest (17.3%). The Ped-CMU ExFPS demonstrated acceptable temporal external validity, supporting its use as a rule-out tool.

Keywords: Extubation failure, Clinical prediction model, External validation, Paediatric cardiac surgery, Paediatric intensive care

Introduction

Mechanical ventilation is a critical component of care in pediatric patients with congenital or acquired heart disease, particularly following cardiac surgery or during episodes of acute cardiorespiratory compromise[1-3]. Although lifesaving, prolonged mechanical ventilation is associated with significant complications, including ventilator-associated pneumonia, diaphragmatic dysfunction, and prolonged intensive care unit stay[4]. Therefore, timely and accurate assessment of readiness for weaning and extubation is essential to optimize patient outcomes and reduce morbidity and healthcare burden [5, 6].

Previous studies conducted in the Pediatric Cardiac Intensive Care Unit (PCICU) at Chiang Mai University have investigated factors associated with extubation failure in this high-risk population. A retrospective study in 2023 identified several significant predictors of extubation failure, including a history of pneumonia, prior re-intubation, post-extubation stridor, and palliative cardiac surgery[3]. Building on these findings, a subsequent study in 2024 developed the Pediatric Chiang Mai University Extubation Failure Predictive Score (Ped-CMU ExFPS) using data collected between 2016 and 2022. The model demonstrated good discriminative ability (area under the receiver operating characteristic curve [AUROC] of 0.77) and satisfactory calibration, incorporating three key predictors: history of pneumonia, prior re-intubation, and physiologic cyanosis with oxygen saturation (SpO_2) greater than 85% [7].

The Ped-CMU ExFPS was designed to support clinical decision-making by identifying pediatric cardiac patients at increased risk of extubation failure. However, external validation of prediction models in independent cohorts is essential before clinical implementation, as model performance may decline when applied to different patient populations and clinical settings. In

particular, variability in case-mix, surgical complexity, and clinical practice patterns may influence the generalizability and real-world performance of the model.

Extubation decisions in pediatric cardiac patients are inherently complex and influenced by multiple physiologic and perioperative factors (8, 9), including whether the patient is in the postoperative period following cardiac surgery. Post-operative patients may have additional risks such as residual sedation, fluid shifts, and respiratory muscle weakness, which may affect extubation outcomes[8].

Therefore, this study aims to externally validate the Ped-CMU ExFPS by evaluating its discrimination, calibration, and clinical applicability in a contemporary cohort of pediatric cardiac patients. In addition, an exploratory analysis will be conducted to assess clinician adherence to the score's recommendations, particularly in cases of extubation failure. This approach will help identify potential gaps between predicted risk and real-world clinical decision-making, as well as opportunities to improve implementation of the score in routine PCICU practice.

Results

Baseline Characteristics

A total of 142 extubation episodes from the validation cohort were analyzed and compared with 352 episodes from the original development cohort. Baseline characteristics of both cohorts are summarized in *Table 1.* Patients in the validation cohort were significantly older (31 ± 41 vs. 20 ± 30 months, $p = 0.004$), with higher body weight (10 ± 10 vs. 8 ± 8 kg, $p = 0.014$) and height (76 ± 27 vs. 69 ± 22 cm, $p = 0.006$). The proportion of males was comparable between groups (56% vs. 52%, $p = 0.371$). Genetic comorbidities were less frequent in the validation cohort (21%

vs. 30%, $p = 0.045$). Notably, the validation cohort demonstrated a significantly lower prevalence of native cyanotic heart disease (30% vs. 50%, $p < 0.001$), ICU lesion-related cyanosis (23% vs. 35%, $p = 0.010$), and pediatric pulmonary arterial hypertension (18% vs. 33%, $p < 0.001$). Oxygen saturation was slightly higher in the validation cohort (95 ± 7% vs. 93 ± 7%, $p = 0.012$). Postoperative status differed significantly between cohorts ($p < 0.001$), with a higher proportion of postoperative patients in the validation dataset. Additionally, pneumonia was less common in the validation cohort (27% vs. 43%, $p < 0.001$), while steroid use prior to extubation was markedly lower (14% vs. 51%, $p < 0.001$). The incidence of extubation failure was numerically lower in the validation cohort (7.7% vs. 11.0%), although this difference did not reach statistical significance ($p = 0.257$).

External Validation

Comparison of Score Distribution and Case-Mix

The membership model yielded an AUROC of 0.72 (95% CI: 0.64–0.80), indicating fair discriminative ability in distinguishing between the development and validation cohorts, and suggesting a moderate degree of dissimilarity in predictor characteristics between the two datasets. Similar, the distribution of the Ped-CMU ExFPS and its corresponding linear predictor were compared between the two datasets. Neither the total score (2.4 ± 3.7 in the development dataset vs. 2.6 ± 3.4 in the validation dataset) nor the linear predictor (−2.73 ± 0.85 vs. −2.68 ± 0.78, respectively) differed significantly between cohorts (both P = 0.570) (Table 2).

Predictor–Outcome Associations

The associations between predictors and extubation failure in both the development and validation datasets are presented in *Table 3*. In the validation cohort, a history of re-intubation remained a strong and statistically significant predictor of extubation failure (OR 12.0, 95% CI 2.53–67.2, $p = 0.002$), with an even larger effect size compared to the development model (OR 6.77). In contrast, pneumonia, which was a significant predictor in the development cohort (OR 4.12, $p < 0.001$), was not associated with extubation failure in the validation dataset (OR 0.83, $p = 0.859$). Cyanosis-related variables could not be reliably estimated in the validation cohort due to sparse events in certain categories, limiting direct comparison.

Model Performance

Discrimination

The model demonstrated acceptable discriminative performance in the validation cohort, with an AUROC of 0.77 (95% CI 0.61–0.93), comparable to the development cohort (AUROC 0.76, 95% CI 0.68–0.85). The AUROC (C-statistic) remained within a clinically useful range, consistent with the original development study (*Figure 2*).

Calibration

Calibration plots demonstrated good agreement between predicted and observed risks (Figure 3). The calibration slope was 1.04 (95% CI: 0.46–1.66) and CITL was −0.11 (95% CI: −0.82–0.50), indicating adequate overall calibration.

Diagnostic Performance

Diagnostic performance of the Ped-CMU ExFPS at the predefined cutoff is shown in *Table 4.* In the validation cohort, the model demonstrated: Sensitivity: 81.82% (95% CI 48.22–97.72), Specificity: 67.18% (95% CI 58.43–75.12), and Accuracy: 68.31% (95% CI 59.98–75.86). The model maintained a high negative predictive value (NPV) of 97.78% (95%CI 92.20 - 99.73), indicating strong ability to rule out extubation failure in low-risk patients. However, the positive predictive value (PPV) remained modest at 17.31% (95%CI 8.23 - 30.33), reflecting the relatively low prevalence of the outcome. Likelihood ratios supported these findings, with a modest positive likelihood ratio and a low negative likelihood ratio, reinforcing the model's utility primarily as a rule-out tool in clinical practice.

Discussion

In this external validation study, the Ped-CMU ExFPS demonstrated acceptable performance with preserved discrimination and meaningful clinical utility in a contemporary pediatric cardiac cohort. Importantly, the model retained its ability to identify patients at low risk of extubation failure, as reflected by a consistently high negative predictive value. However, variations in predictor effects and baseline characteristics between the development and validation cohorts underscore the dynamic and context-dependent nature of extubation risk in pediatric cardiac intensive care [1, 4, 9].

Physiological Interpretation of Key Findings

Among the evaluated predictors, a history of re-intubation emerged as the most robust and consistent determinant of extubation failure. Re-intubation can cause mucosal damage through

mechanical shear stress, resulting in airway trauma and edema, thereby predisposing patients to post-extubation stridor (PES). Furthermore, repeated intubation contributes to prolonged cumulative ventilation time in the PCICU. Previous studies have consistently demonstrated that prolonged intubation is associated with an increased risk of extubation failure [10, 11]. In pediatric cardiac patients, these factors are often exacerbated by residual cardiac lesions, ventricular dysfunction, and abnormal cardiopulmonary interactions, all of which increase the likelihood of weaning failure [7].

In contrast, pneumonia was not significantly associated with extubation failure in the validation cohort. This may be related to differences in case-mix or limited statistical power, as reflected by the wide confidence interval, rather than a true absence of effect. This discrepancy may reflect temporal improvements in infection control, earlier diagnosis, and optimized peri-extubation management. While pneumonia impairs gas exchange and increases the work of breathing, its impact may be mitigated when adequately treated before extubation, thereby reducing its predictive value in contemporary practice [12].

Although cyanosis-related variables could not be estimated in the validation cohort due to the sparse number of events, the underlying physiological rationale remains clinically relevant. Patients with cyanotic heart disease, particularly those with imbalanced pulmonary-to-systemic blood flow (Qp/Qs mismatch), are highly sensitive to changes in pulmonary vascular resistance and oxygen delivery. Following extubation, the withdrawal of positive pressure ventilation may exacerbate these imbalances, predisposing to respiratory and hemodynamic instability [3, 7, 13, 14].

Impact of Case-Mix and Temporal Changes

Despite similar score distributions and linear predictor across cohorts, several differences in baseline characteristics were observed, including a lower prevalence of pulmonary hypertension and pneumonia. Notably, the lower use of prophylactic steroids in the validation cohort may also reflect evolving clinical practice. As corticosteroids are known to attenuate airway inflammation and reduce the risk of post-extubation stridor, variation in their use may have contributed to differences in airway-related outcomes and extubation success between cohorts. These findings suggest evolving clinical practice and improvements in perioperative and critical care management.

Prediction models are inherently sensitive to case-mix variation, and performance may decline when applied to populations with different baseline risks or treatment patterns [15, 16].These findings reinforce the need for external validation and iterative model updating to maintain clinical relevance over time. They also underscore the importance of integrating prediction models into adaptive clinical pathways, rather than relying on static risk estimates.

Clinical Implications and Implementation

From a clinical perspective, the Ped-CMU ExFPS appears most useful as a risk stratification and rule-out tool. Importantly, translating predictive models into clinical benefit requires effective implementation. Extubation decisions are complex and influenced by clinician judgment, institutional protocols, and patient-specific factors not captured by prediction tools [8]. Therefore, integrating the Ped-CMU ExFPS into structured clinical workflows, such as

standardized extubation protocols or electronic decision-support systems, may enhance adherence and improve the consistency of care.

Future Directions

Future studies should focus on multicenter validation to enhance generalizability. Additionally, incorporation of dynamic clinical variables, including respiratory muscle strength, sedation level, and post-extubation airway factors such as stridor, may further improve predictive accuracy.

Strengths and Limitations

This study has several strengths, including temporal external validation, adherence to methodological standards for prediction model evaluation, and comprehensive assessment of discrimination and calibration (13). This study also has several limitations. First, the small sample size of the validation dataset led to a low event prevalence in certain categories of the cyanosis predictor, limiting reliable estimation of the corresponding regression coefficient. Second, as the study was conducted at a single tertiary care center, the findings may not be generalizable to other settings. Although external validation demonstrated good reproducibility within our institution, further validation in geographically and institutionally diverse populations with larger sample sizes is warranted to assess model transportability. Third, subgroup and algorithmic fairness analyses were not performed owing to the small number of outcome events. This should be addressed in larger multicentre validation.

Conclusion

The Ped-CMU ExFPS showed acceptable discrimination, calibration, and diagnostic indices in external validation. Its high sensitivity supports its role as a safe screening tool for identifying low-risk candidates for extubation. Given that this represents a temporal external validation within the same institution, the score appears well-suited to the local Thai pediatric cardiac population at Chiang Mai University, though broader geographic validation is warranted before wider adoption.

Methods

Study Design and Setting

This retrospective observational cohort study was conducted in the Pediatric Cardiac Intensive Care Unit (PCICU) at the Faculty of Medicine, Chiang Mai University. Medical records of all eligible extubation events between June 1, 2022, and June 30, 2024, were reviewed. The overall study workflow is summarized in Figure 1.

Each extubation attempt was considered as an independent unit of analysis (extubation episode). Therefore, individual patients could contribute more than one observation if multiple extubation attempts occurred during the same or different hospital admissions. Two datasets were defined: The development and validation datasets were derived from the same clinical setting, with identical eligibility criteria, outcome definition, and candidate predictors. The datasets differed only in the period of participant recruitment; the development dataset was collected from July 2016 to May 2022 [7], whereas the validation dataset was collected from June 2022 to June 2024. Given these characteristics, this study constitutes a temporal external validation.

The study protocol was reviewed and approved by the Human Research Ethics Committee (Panel 5), Faculty of Medicine, Chiang Mai University (reference number 147/2026; study code PED-2569-0068), and was performed in accordance with the principles of the Declaration of Helsinki. Given the retrospective chart review design and minimal risk to participants, the committee granted a waiver of written informed consent.

Study Population

Eligible participants were pediatric patients aged 1 month to 18 years with either congenital or acquired heart disease who required invasive mechanical ventilation and underwent at least one extubation attempt during the study period. Each documented extubation attempt between June 1, 2022, and June 30, 2024, was considered an independent observation and analyzed as a distinct extubation episode. Patients were excluded if they had a tracheostomy in place at the time of

extubation, experienced unplanned extubation events (e.g., self-extubation), or were discharged or died while still intubated. In addition, extubation episodes with incomplete core data that precluded accurate outcome ascertainment were excluded from the analysis.

Weaning and Extubation Protocol

Weaning from mechanical ventilation was initiated once patients achieved stable hemodynamic status and adequate oxygenation. Patients underwent a 30-minute spontaneous breathing trial (SBT) using either synchronized intermittent mandatory ventilation (SIMV) with pressure support or continuous positive airway pressure (CPAP), according to the unit protocol.

Extubation was performed when patients met all clinical readiness criteria based on the institutional Weaning and Extubation Assessment Form. This assessment included the Pediatric Chiang Mai University Extubation Failure Predictive Score (Ped-CMU ExFPS), which comprises history of pneumonia, prior re-intubation, and physiologic cyanosis.

Data Collection

Data were collected from medical records and bedside documentation, including demographic and clinical characteristics (age, sex, primary diagnosis, type of cardiac surgery, and post-operative status) and pre-extubation assessment results, including the Ped-CMU Extubation Failure Predictive Score (Ped-CMU ExFPS). Post-extubation outcomes were assessed within 48 hours following each extubation attempt. Post-operative status was defined as patients who had undergone cardiac surgery during the same hospital admission prior to the extubation attempt.

Definition

- Extubation failure was defined as the need for re-intubation within 48 hours after a specific extubation attempt due to respiratory failure or clinical deterioration[17]. Re-intubation occurring beyond 48 hours was not considered extubation failure for that episode; however, such cases could be included as a new extubation episode if followed by another extubation attempt.
- Physiologic cyanosis was defined according to the patient's underlying cardiovascular physiology at the time of intubation. For example, in patients with Tetralogy of Fallot (TOF) status post modified right Blalock–Thomas–Taussig (mRMBT) shunt, oxygen saturation levels below 95% were considered physiologically appropriate [3].
- Physiologic acyanosis was defined according to the patient's underlying cardiovascular physiology at the time of intubation. For example, in patients with ventricular septal defect (VSD) or Tetralogy of Fallot (TOF) status post total correction, oxygen saturation levels ≥95% were considered physiologically appropriate [3].
- History of re-intubation was defined as a re-intubation event during PCICU admission[3].
- Pneumonia was defined based on clinical, radiographic, and microbiological criteria, including fever, productive sputum, pulmonary infiltrates on chest X-ray (CXR), and/or a positive sputum culture prior to extubation. A history of pneumonia encompassed all etiologies, including community-acquired pneumonia (CAP), hospital-acquired pneumonia (HAP), and ventilator-associated pneumonia (VAP) [3].

Study Endpoint

The study endpoint of external validation was reintubation event (yes/no) recorded in electronic medical record (EMR).

Statistical Analysis and Sample Size

Continuous variables are expressed as mean ± standard deviation (SD) for normally distributed data or median (interquartile range, IQR) for non-normally distributed data. Categorical variables are presented as frequency (percentage). Comparisons between groups were performed using Student's t-test for normally distributed continuous variables, the Mann–Whitney U test for non-normally distributed continuous variables, and the chi-square test or Fisher's exact test for categorical variables. A two-sided P-value <0.05 was considered statistically significant. All statistical analyses were performed using R version 4.5.2 (R Foundation for Statistical Computing, Vienna, Austria).

Sample size requirements were estimated using the *pmvalsampsize* package, with key parameters derived from the original model development study. To meet Criterion 2 (precise estimation of the calibration slope), a sample of 303 participants was required. However, due to time constraints and low case prevalence, all eligible participants from the database (n=142) were included to maximize statistical power and ensure generalizability of findings.[18]

Sample size power was evaluated using the approach proposed by Riley et al. Despite the smaller-than-optimal sample size, our cohort of 142 participants provided a 95% confidence interval (CI) width of 0.884 for the observed-to-expected (O/E) ratio (assuming an O/E ratio of 1.0), a 95% CI width of 0.73 for the calibration slope, and a 95% CI width of 0.219 for the area

under the receiver operating characteristic curve (AuROC) (assuming an AuROC of 0.70). *Supplementary Figures 1–3* illustrate the expected precision across various plausible values of these performance indices.[18]

This study employed retrospective data collection. The proportion and pattern of missing data were assessed for all predictor variables. For missing data that exceeded 5%, multiple imputation was performed using chained equations.

External Validation

The validation analysis followed the framework proposed by Debray et al. First, baseline characteristics, including all predictors and outcomes, were compared between the development and validation datasets using appropriate statistical tests as described above. Second, a membership model was fitted to assess the degree of dissimilarity between the two cohorts through discrimination analysis. In the context of external validation, the development and validation cohorts are expected to differ in participant characteristics, encompassing both predictor variables and outcomes, as a low degree of overlap between the two cohorts supports the generalizability of the validation findings. The degree of dissimilarity between the two datasets was quantified using the membership model, in which an AUROC of 0.5 indicates that the two cohorts are indistinguishable with respect to participant characteristics and outcome distribution, whereas an AUROC approaching 1.0 indicates substantial dissimilarity. Third, the distribution of the Ped-CMU ExFPS linear predictor, expressed as mean ± SD, was compared between the development and validation datasets using an independent samples Student's t-test to evaluate case-mix heterogeneity[15].

Predictor-outcome associations in the validation dataset were analyzed using multivariable logistic regression with the same predictors specified in the original model development study.[9]

Model performance assessment

Model performance in the validation dataset was assessed through discrimination and calibration measures. Discrimination was quantified using area under the receiver operating characteristic curve (AUROC), with 95% confidence intervals estimated. The AUROC represents the probability that a randomly selected individual with the outcome receives a higher predicted probability than a randomly selected individual without the outcome, ranging from 0.5 (no discrimination) to 1.0 (perfect discrimination). An AUROC between 0.70 and 0.80 was considered acceptable[19].

Calibration was assessed using the calibration-in-the-large (CITL) and calibration slope, presented graphically via calibration plots. The CITL reflects overall correspondence between mean predicted probability and observed outcome frequency (ideal value = 0), where negative values indicate overestimation and positive values indicate underestimation. The calibration slope reflects the spread of predicted probabilities across the prediction range (ideal value = 1), where values above 1 indicate insufficient variation and values below 1 indicate excessive variation in predictions.

Model updating

In the case of poor calibration, model updating was considered using two strategies. Recalibration intercept, which addresses differences in baseline risk by re-estimating the model intercept, was applied to models with poor CITL. Recalibration slope, which recalibrates the linear predictor of the original model, was applied to models with poor calibration slope and/or CITL.

Model extension was not considered, as no new predictors were introduced in the validation dataset [20]. In the event of poor discrimination, potential handling method would include model refitting and reweighting through the development of a new prediction model with the same predictors or model extension by incorporating additional meaningful predictors or biomarkers [20].

Informed consent

Patient consent was waived due to the retrospective design of the study. The research involved the analysis of deidentified secondary data from admission records, posed minimal risk to the participants, and did not affect patient care.

Declaration of AI Use

The authors used an AI language model (Claude, Anthropic, PBC) to assist with English language editing and grammar improvement during manuscript preparation. The AI tool was not used for data analysis, interpretation of results, or generation of scientific content. The authors take full responsibility for the integrity and accuracy of all reported findings.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Code Availability

The analytical code used for the external validation is openly available at https://github.com/Nopkhongthon/ped-cmu-exfps-validation.

References

1. Farias JA, Alía I, Retta A, Olazarri F, Fernández A, Esteban A, et al. An evaluation of extubation failure predictors in mechanically ventilated infants and children. Intensive Care Med. 2002;28(6):752–7.
2. Li Y, Jia Y, Wang H, Wu X, Li S, Yan F, et al. Early extubation is associated with improved outcomes after complete surgical repair of pulmonary atresia with ventricular septal defect and hypoplastic pulmonary arteries in pediatric patients. J Cardiothorac Surg. 2021;16(1):31.
3. Saengsin K, Sittiwangkul R, Borisuthipandit T, Trongtrakul K, Tanasombatkul K, Phanacharoensawad T, et al. Predictive factors of extubation failure in pediatric cardiac intensive care unit: A single-center retrospective study from Thailand. Front Pediatr. 2023;11:1156263.
4. Zhao QY, Wang H, Luo JC, Luo MH, Liu LP, Yu SJ, et al. Development and Validation of a Machine-Learning Model for Prediction of Extubation Failure in Intensive Care Units. Front Med (Lausanne). 2021;8:676343.
5. Baptistella AR, Mantelli LM, Matte L, Carvalho M, Fortunatti JA, Costa IZ, et al. Prediction of extubation outcome in mechanically ventilated patients: Development and validation of the Extubation Predictive Score (ExPreS). PLoS One. 2021;16(3):e0248868.
6. Elisa P, Francesca C, Marco P, Davide V, Laura Z, Fabrizio Z, et al. Ventilation Weaning and Extubation Readiness in Children in Pediatric Intensive Care Unit: A Review. Front Pediatr. 2022;10:867739.
7. Saengsin K, Sittiwangkul R, Borisuthipandit T, Wongyikul P, Tanasombatkul K, Phanacharoensawad T, et al. Development of a clinical prediction tool for extubation failure in pediatric cardiac intensive care unit. Front Pediatr. 2024;12:1346198.
8. Melton N, Lazar JF, Childers WK, Anderson D, Jaik NP, Loran DB, et al. Preventing Respiratory Failure After Cardiac Surgery Using Post-extubation Bilevel Positive Airway Pressure Therapy. Cureus. 2019;11(3):e4236.
9. Saengsin K, Sittiwangkul R, Borisuthipandit T, Wongyikul P, Tanasombatkul K, Phanacharoensawad T, et al. Development of a clinical prediction tool for extubation failure in pediatric cardiac intensive care unit. Frontiers in Pediatrics. 2024;Volume 12 - 2024.
10. Silva-Cruz AL, Velarde-Jacay K, Carreazo NY, Escalante-Kanashiro R. Risk factors for extubation failure in the intensive care unit. Rev Bras Ter Intensiva. 2018;30(3):294–300.
11. Gupta P, Rettiganti M, Gossett JM, Yeh JC, Jeffries HE, Rice TB, et al. Risk factors for mechanical ventilation and reintubation after pediatric heart surgery. J Thorac Cardiovasc Surg. 2016;151(2):451–8.e3.
12. Driver C. Pneumonia part 1: Pathology, presentation and prevention. Br J Nurs. 2012;21(2):103–6.

13. Ley S, Puderbach M, Risse F, Ley-Zaporozhan J, Eichinger M, Takenaka D, et al. Impact of oxygen inhalation on the pulmonary circulation: assessment by magnetic resonance (MR)-perfusion and MR-flow measurements. Invest Radiol. 2007;42(5):283–90.
14. Barnea O, Austin EH, Richman B, Santamore WP. Balancing the circulation: theoretic optimization of pulmonary/systemic flow ratio in hypoplastic left heart syndrome. J Am Coll Cardiol. 1994;24(5):1376–81.
15. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. Journal of Clinical Epidemiology. 2015;68(3):279–89.
16. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. BMC Medical Research Methodology. 2022;22(1):316.
17. Khemani RG, Markovitz BP, Curley MAQ. Characteristics of children intubated and mechanically ventilated in 16 PICUs. Chest. 2009;136(3):765–71.
18. Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Statistics in Medicine. 2021;40(19):4230–51.
19. Assessing the Fit of the Model. Applied Logistic Regression2000. p. 143–202.
20. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. BMC Med Res Methodol. 2022;22(1):316.

Acknowledgements. This study was partially supported by the Faculty of Medicine, Chiang Mai University.

Author contributions. K.S. conceived and designed the study. T.K. collected the data. N.K. and P.W. performed the statistical analysis. P.P. supervised the methodology. T.K. and N.K. drafted the manuscript. K.S., P.W. and P.P. critically revised the manuscript. All authors reviewed and approved the final version. T.K. and N.K. contributed equally to this work.

Additional Information: All information were provided in manuscript.

Competing interests. The authors declare that they have no competing interests.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Patient and public involvement. No patients or members of the public were involved in the design, conduct, reporting, or dissemination of this study.

Figure 1: Study workflow of the external validation of the Pediatric Chiang Mai University Extubation Failure Predictive Score (Ped-CMU ExFPS).

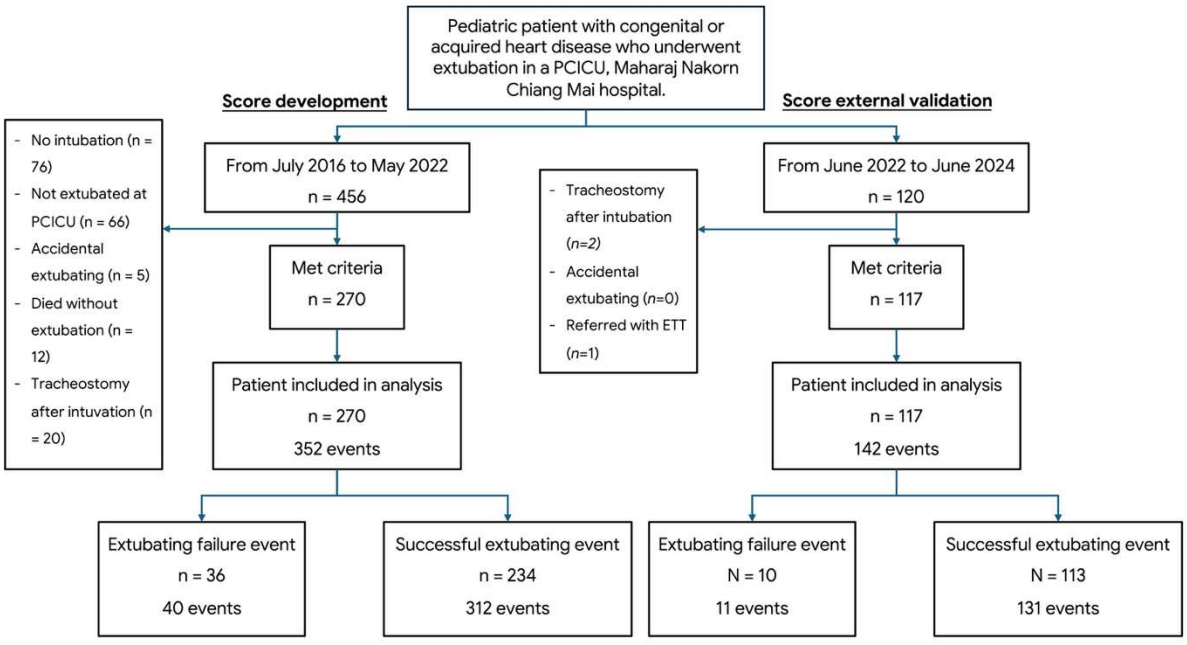


Figure 2: Comparison of ROC curve between development and validation study.

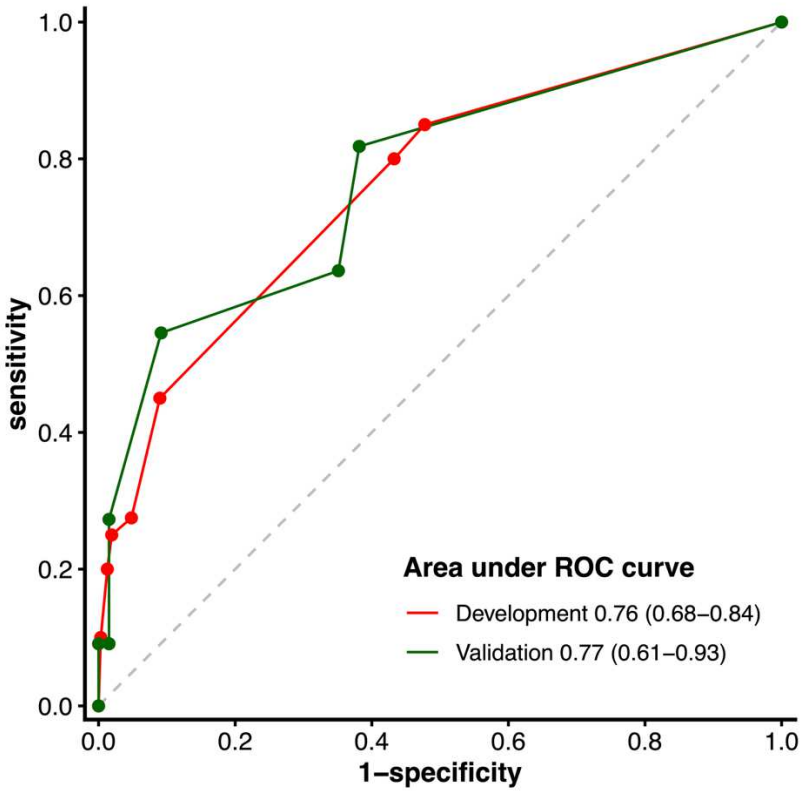


Figure 3: Calibration plot.

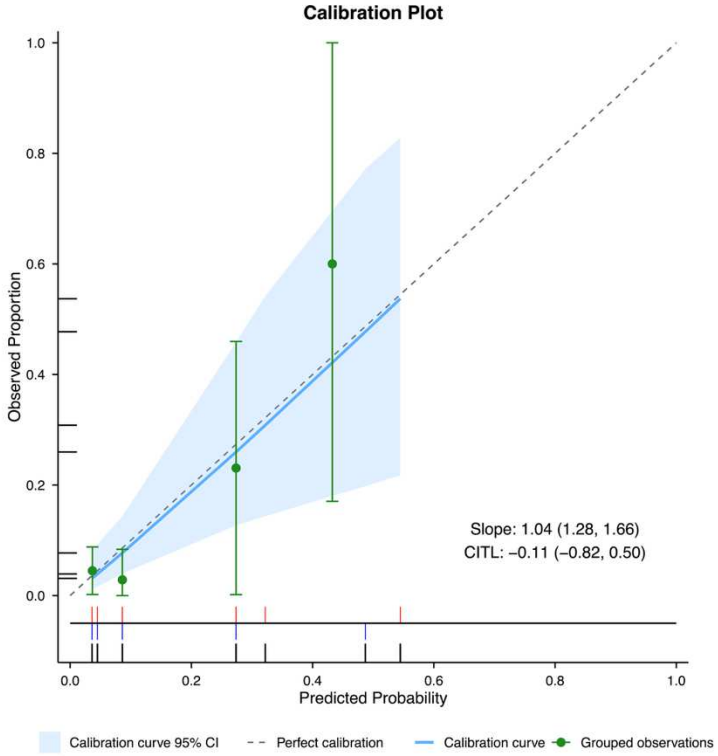


Table 1: Comparison of clinical characteristics of the included patients between the development and validation study

Characteristic	Missing	External validation dataset N = 142[1]	Development dataset N = 352[1]	p-value[2]
Age (month)	0 (0%)	31 (± 41)	20 (± 30)	0.004
Male	0 (0%)	80 (56%)	182 (52%)	0.371
Body weight (kg)	0 (0%)	10 (± 10)	8 (± 8)	0.014
Height (cm)	0 (0%)	76 (± 27)	69 (± 22)	0.006
Body mass index (kg/m2)	0 (0%)	13.85 (± 2.71)	14.14 (± 3.34)	0.316
Having genetic disease	0 (0%)	30 (21%)	107 (30%)	0.045
Genetic disease	0 (0%)			0.130
Down's syndrome		15 (50%)	47 (44%)	
Heterotaxy syndrome		3 (10%)	25 (23%)	
Digeorge's syndrome		6 (20%)	13 (12%)	
VacTERL syndrome		3 (10%)	3 (2.8%)	
Other		3 (10%)	19 (18%)	
Cardiac disease	0 (0%)			
Single ventricle		0 (0%)	62 (18%)	
VSD		40 (28%)	77 (22%)	
PDA		0 (0%)	19 (5.4%)	
COA		4 (2.8%)	13 (3.7%)	
ASD		10 (7.0%)	11 (3.1%)	
Hemitruncus / truncus		3 (2.1%)	10 (2.8%)	
IAA		1 (0.7%)	1 (0.3%)	
TOF		29 (20%)	53 (15%)	
DORV		14 (9.9%)	17 (4.8%)	
TGA		6 (4.2%)	24 (6.8%)	
ccTGA		4 (2.8%)	6 (1.7%)	
AVSD		3 (2.1%)	14 (4.0%)	
Abnormal coro / ALCAPA		1 (0.7%)	1 (0.3%)	
PS		5 (3.5%)	8 (2.3%)	
Ebstein		0 (0%)	0 (0%)	
Hypertropic		1 (0.7%)	4 (1.1%)	
TAPVR		0 (0%)	5 (1.4%)	
Other		21 (15%)	27 (7.7%)	
Native anatomy cyanosis	0 (0%)	43 (30%)	175 (50%)	<0.001
Status ICU Lesion Cyanosis	0 (0%)	32 (23%)	122 (35%)	0.010
Oxygen saturation (%)	2 (0.4%)	95 (± 7)	93 (± 7)	0.012

Characteristic	Missing	External validation dataset N = 142[1]	Development dataset N = 352[1]	p-value[2]
Paediatric pulmonary arterial hypertension	0 (0%)	26 (18%)	117 (33%)	<0.001
Duration of intubation (days)	0 (0%)	10 (± 20)	9 (± 10)	0.468
Post-operation	0 (0%)			<0.001
No operation		26 (18%)	140 (40%)	
Palliative surgery		116 (82%)	208 (59%)	
Total correction		0 (0%)	4 (1.1%)	
Pneumonia	0 (0%)	38 (27%)	153 (43%)	<0.001
History of intubation	0 (0%)	18 (13%)	26 (7.4%)	0.080
Treated with steroid	0 (0%)	20 (14%)	178 (51%)	<0.001
Balance per kg	9 (1.8%)	5 (± 112)	7 (± 34)	0.811
Extubation failure	0 (0%)			0.257
Fail		11 (7.7%)	40 (11%)	
Success		131 (92%)	312 (89%)	
Death	0 (0%)	3 (2.1%)	13 (3.7%)	0.575

[1]Mean (± SD); n (%)

[2]Welch Two Sample t-test; Fisher's exact test; NA

Abbreviations: ALCAPA = anomalous left coronary artery from the pulmonary artery; ASD = atrial septal defect; AVSD = atrioventricular septal defect; BMI = body mass index; ccTGA = congenitally corrected transposition of the great arteries; COA = coarctation of the aorta; DORV = double outlet right ventricle; IAA = interrupted aortic arch; PDA = patent ductus arteriosus; PS = pulmonary stenosis; SD = standard deviation; TAPVR = total anomalous pulmonary venous return; TGA = transposition of the great arteries; TOF = tetralogy of Fallot; VACTERL = vertebral anomalies, anal atresia, cardiac defects, tracheoesophageal fistula, renal anomalies, and limb abnormalities; VSD = ventricular septal defect.

[1] Mean (±SD); n(%)

[2] Two-sample t-test; Fisher's exact test

Table 2: Comparison of original score and linear predictor of score between development and validation study

Characteristic	External validation dataset N = 142[1]	Development dataset N = 352[1]	p-value[2]
Ped-CMU ExFPS score	2.4 (± 3.7)	2.6 (± 3.4)	0.570
Linear predictor of Ped-CMU ExFPS	-2.73 (± 0.85)	-2.68 (± 0.78)	0.570

[1]Mean (± SD)

[2]Welch Two Sample t-test

Table 3: The predictor-outcome associations in the development and the validation study

	Development model				Validation model			
Characteristic	**Beta**	**OR**	**95% CI**	**p-value**	**Beta**	**OR**	**95% CI**	**p-value**
History of reintubation	1.912	6.77	2.51, 18.1	<0.001	2.489	12.0	2.53, 67.2	0.002
Pneumonia status	1.417	4.12	1.93, 9.48	<0.001	-0.186	0.83	0.08, 5.65	0.859
Cyanosis status[1]								
Cyanosis (SpO2 ≤ 85%)	Ref				Ref			
Acyanosis	0.218	1.24	0.48, 3.65	0.668	NA	NA	NA	NA
Cyanosis (SpO2 > 85%)	1.584	4.87	1.67, 15.7	0.005	NA	NA	NA	NA

[1]Cyanosis estimates could not be calculated in the validation model due to low prevalence events in one category

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

Table 4: Diagnostic indices

Statistic	Development[1]	Validation[1]
Prevalence	11.36% (8.24, 15.15)	7.75% (3.93, 13.44)
Sensitivity	80.00% (64.35, 90.95)	81.82% (48.22, 97.72)
Specificity	61.54% (55.89, 66.96)	67.18% (58.43, 75.12)
Accuracy	63.64% (58.37, 68.67)	68.31% (59.98, 75.86)
PPV	21.05% (14.87, 28.40)	17.31% (8.23, 30.33)
NPV	96.00% (92.27, 98.26)	97.78% (92.20, 99.73)
LR positive	2.08 (1.68, 2.56)	2.49 (1.72, 3.61)
LR negative	0.32 (0.17, 0.60)	0.27 (0.07, 0.95)

[1] Proportion (95%CI)

Abbreviations: CI = confidence interval, LR = likelihood ratio; NPV = negative predictive value; PPV = positive predictive value

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplement260610V4clear.docx