J Korean Med Sci. 2025 Mar 17;40(10):e106
https://doi.org/10.3346/jkms.2025.40.e106
eISSN 1598-6357·pISSN 1011-8934

JKMS

Brief Communication
Pediatrics

Validation of the Phoenix Criteria for Sepsis and Septic Shock in a Pediatric Intensive Care Unit

Chang Hoon Han [1,2] **Hamin Kim** [3] **Mireu Park** [3] **Soo Yeon Kim** [3] **Jong Deok Kim** [3] **Myung Hyun Sohn** [3] **Seng Chan You** [1,2] **and Kyung Won Kim** [2,3]

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea
[2]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Korea
[3]Department of Pediatrics, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

Received: Aug 28, 2024
Accepted: Feb 9, 2025
Published online: Mar 4, 2025

Address for Correspondence:
Kyung Won Kim, MD, PhD
Department of Pediatrics, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.
Email: kwkim@yuhs.ac

Seng Chan You, MD, PhD
Department of Biomedical Systems Informatics, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.
Email: chandryou@yuhs.ac

ORCID iDs
Chang Hoon Han
https://orcid.org/0000-0002-8092-5884
Hamin Kim
https://orcid.org/0000-0002-0601-7300
Mireu Park
https://orcid.org/0000-0003-4342-6143
Soo Yeon Kim
https://orcid.org/0000-0003-4965-6193
Jong Deok Kim
https://orcid.org/0000-0002-4266-5655

ABSTRACT

The applicability of the Phoenix criteria and Phoenix Sepsis Score in higher-resource pediatric intensive care units (PICUs) outside the United States requires further validation. A retrospective cohort study analyzed electronic health records of 1,304 PICU admissions under 18 years old with suspected infection between February 2017 and December 2023. The score was calculated using two methods: 24-hour assessment, based on worst sub-scores within 24 hours of admission, and prompt assessment, using values closest to admission within 6 hours before or after. Based on the 24-hour assessment, in-hospital mortality was 8.3% for sepsis and 10.3% for septic shock. The score demonstrated an area under the precision-recall curve of 0.42 (95% confidence interval, 0.31–0.55) for in-hospital mortality. Results were consistent across both assessment methods. The Phoenix criteria and the Phoenix Sepsis Score are reliable predictors of mortality outcomes. Further investigation in diverse clinical settings is warranted.

Keywords: Sepsis; Septic Shock; Pediatrics; Intensive Care Units; Mortality; Predictive Value of Tests

The Phoenix criteria, developed by the Society of Critical Care Medicine Pediatric Sepsis Definition Task Force, aims to replace outdated criteria with a scoring system that better correlates with mortality across diverse healthcare settings.[1] Despite pediatric sepsis remaining a global healthcare challenge, a lack of updated criteria based on large-scale data persists.[2-4] The Phoenix sepsis criteria, defining sepsis as a suspected infection with life-threatening organ dysfunction, and the Phoenix Sepsis Score, an advanced organ dysfunction scoring system developed specifically for this purpose, appear to address this unmet need.[5]

While the Phoenix sepsis score has demonstrated competent performance in both higher- and lower-resource settings, its validation in higher-resource settings have been limited to centers in the United States, and further evaluation in other countries is warranted. Moreover, although the current method utilizes the worst sub-scores assessed within a 24 hours window to ensure sensitive detection of any organ dysfunction in that period, its usage is limited for representing the acute status of a patient.[1,5,6]

Myung Hyun Sohn
https://orcid.org/0000-0002-2478-487X
Seng Chan You
https://orcid.org/0000-0002-5052-6399
Kyung Won Kim
https://orcid.org/0000-0003-4529-6135

Funding

This research was supported by a grant of the MD-PhD/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C1010462).

Disclosure

Seng Chan You reports being a CEO of PHI Digital Healthcare; and grants from Daiichi Sankyo. The other authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Han CH, Kim H, You SC, Kim KW. Data curation: Han CH. Formal analysis: Han CH. Investigation: Han CH, Kim H, You SC, Kim KW. Methodology: Han CH, Kim H, You SC, Kim KW. Supervision: You SC, Kim KW. Validation: Han CH, Kim H, You SC, Kim KW. Writing - original draft preparation: Han CH. Writing - review & editing: Han CH, Kim H, Park M, Kim SY, Kim JD, Sohn MH, You SC, Kim KW.

This study aimed to validate the Phoenix criteria in a higher-resource pediatric intensive care unit (PICU) setting in Korea. We also sought to explore their applicability in more acute settings, with a time window shorter than 24 hours for data collection.

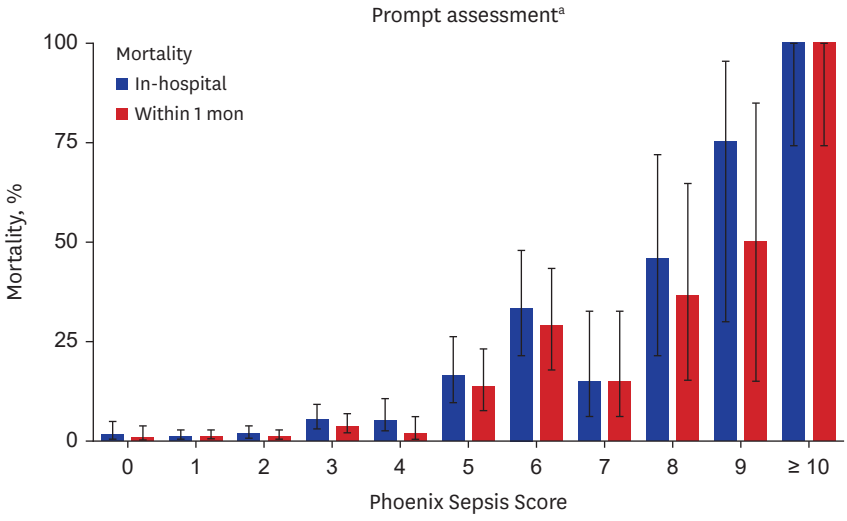
A single-center retrospective cohort study was conducted using electronic health records (EHRs) data of children admitted to the PICU at Severance Hospital (Seoul, Korea). We included PICU admissions of children under 18 years of age between February 2017 and December 2023. Each encounter was treated separately, and PICU stays < 24 hours were excluded. This study was reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement.[7]

The primary outcome was in-hospital mortality. Mortality within 1 month after PICU admission was also considered as a secondary outcome. Suspected infection was identified by administration of systemic antibacterials or antifungals with culture studies on bacteria or fungi, within 24 hours after admission to the PICU.[5] For each PICU admission with suspected infection, we calculated the Phoenix Sepsis Score using two methods.[1,5] First, "24-hour assessment" was defined as the score calculated from the worst value for each sub-score, determined using data collected within 24 hours after transfer to the unit. Second, to test the score's applicability in practical clinical settings, we defined "prompt assessment" as the score using the values recorded closet to the admission time, within a window of 6 h before and after admission to the PICU. This approach was designed to reflect routine clinical practice where rapid decision-making is often required based on limited initial data. Sensitivity, specificity, and positive predictive values (PPVs) of the Phoenix criteria derived from each method for the mortality outcomes were calculated. We also compared the Phoenix Sepsis Score with the Phoenix-8 score and the pediatric Sequential Organ Failure Assessment (pSOFA) score, calculated using the same methods.[5]

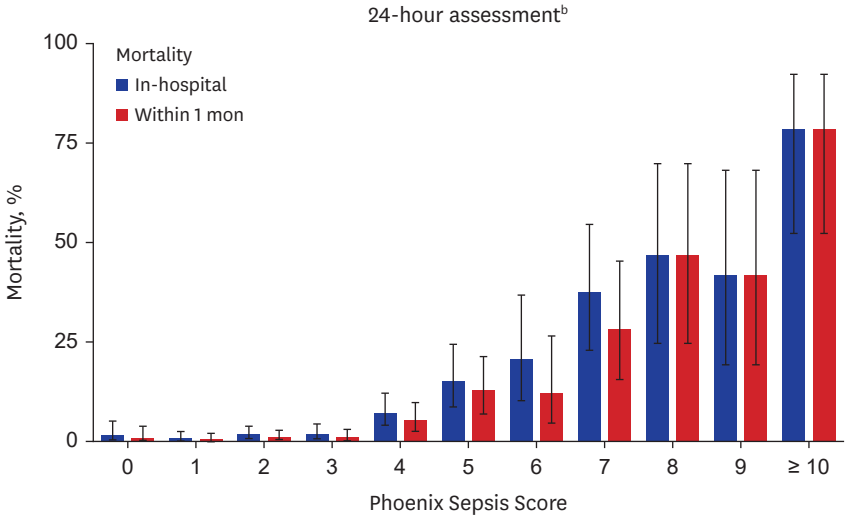
Wilson intervals with continuity correction were used to represent 95% confidence intervals of binomial outcomes.[8] The performance of each score was evaluated using both the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). A high AUROC indicates strong discriminative ability across all thresholds, distinguishing between positive and negative cases. On the other hand, a high AUPRC reflects better precision and recall and signifies better performance in identifying true positives, particularly in tasks involving imbalanced class distributions, such as predicting mortality in pediatric patients.[9] All analyses were performed using R, v.4.1.3 (R Foundation for Statistical Computing, Vienna, Austria).

A total of 901 patients and 1,304 PICU admissions with suspected infection was included in the final analysis (**Supplementary Fig. 1**). Patient demographics are detailed in **Supplementary Table 1**. There were 78 inpatient deaths, with an in-hospital mortality rate of 6.0%. Of these, 65 died within one month of PICU transfer, resulting in a one-month mortality rate of 5.0%. Based on the prompt assessment, 708 cases (54.3%) were classified as sepsis, with 489 cases (37.5%) meeting the criteria for septic shock. Under the 24-hour assessment, 887 cases (68.0%) were identified as sepsis, with 609 cases (46.7%) classified as septic shock. The Phoenix criteria for sepsis exhibited sensitivity of 93.6% to 96.7%, specificity of 33.4% to 48.2%, and PPV of 6.7% to 10.3% across mortality outcomes and assessment methods. For septic shock, it showed sensitivity of 71.8% to 80.8%, specificity of 55.0% to 64.7%, and PPV of 8.1% to 11.5% (**Supplementary Table 2**). The numbers of encounters with outcomes and mortality rates in each Phoenix Sepsis Score category

are shown in the **Fig. 1**, with the prompt assessment and 24-hour assessment displayed separately. In both methods, a general trend of increasing mortality was observed with higher score values. For scores of 8 or greater, mortality outcomes were poorer when the prompt assessment was used, suggesting that fewer survivors were assigned higher scores with this method (**Fig. 1**).



No. of events											
Admissions	81	173	211	123	82	30	31	13	6	4	5
In-hospital mortality	0	1	9	7	12	11	6	6	5	3	5
Mortality within 1 mon	0	0	7	4	9	9	5	5	5	3	5



No. of events											
Admissions	140	285	302	233	157	80	34	32	15	12	14
In-hospital mortality	2	2	5	4	11	12	7	12	7	5	11
Mortality within 1 mon	1	1	3	2	8	10	4	9	7	5	11

Fig. 1. Bar plots illustrating the calibration of the Phoenix Sepsis Score. The Phoenix Sepsis Score was evaluated using 2 different methods, displayed separately. For each possible value of the Phoenix Sepsis Score, the numbers of cases with suspected infection, in-hospital mortality, and mortality within 1 month are shown. The plots also include Wilson intervals of the binomial distribution (whiskers) for each category.
PICU = pediatric intensive care unit.
[a]Representing scores calculated from sub-scores based on values closest to the PICU admission time.
[b]Representing scores calculated from the worst sub-scores based on values collected within 24 hours after PICU admission.

https://doi.org/10.3346/jkms.2025.40.e106

The Phoenix Sepsis Score had AUPRCs of 0.43 (95% confidence interval [CI], 0.32–0.53) on prompt assessment, and 0.42 (95% CI, 0.31–0.55) with 24-hour assessment. AUPRCs were higher for the one-month mortality outcome, with 0.72 (95% CI, 0.60–0.81) for the prompt assessment and 0.72 (95% CI, 0.60–0.82) for the 24-hour assessment. As shown in **Fig. 2**, these results were comparable to those of the Phoenix-8 Score and superior to pSOFA scores. The AUROCs of the Phoenix Sepsis Score ranged from 0.85 (95% CI, 0.80–0.90) to 0.89 (95% CI, 0.84–0.94) across outcomes and methods, which were comparable to both the Phoenix-8 Score and the pSOFA score (**Fig. 2**). Given the low mortality rate, the higher AUPRC values suggest that the Phoenix Sepsis Score along with the Phoenix-8 Score demonstrate superior performance.[9]

This study validated the Pheonix criteria and the Pheonix Sepsis Score in a higher-resource PICU setting in Korea. The mortality rate among the patients diagnosed with sepsis according to the Phoenix criteria, based on 24-hour assessment, ranged from 6.7% to 8.3%. The AUPRC of the Phoenix Sepsis Score was notably higher than pSOFA score, demonstrating superior performance in mortality prediction, and was similar to the Phoenix-8 Score, which is consistent with the previous finding.[5]

The Phoenix Sepsis Score's superior performance in predicting mortality can be attributed to its development process. Unlike the pSOFA score, which was adapted from the adult SOFA score with age-specific adjustments, the Phoenix Sepsis Score was built using a comprehensive, data-driven approach, by integrating components from multiple organ dysfunction scores, including pSOFA, to optimize predictive accuracy.[5,10] Moreover, while pSOFA has primarily been validated using AUROC, the Phoenix Sepsis Score was designed to maximize AUPRC, ensuring better identification of relatively rare mortality outcomes.[9]

AUPRC

Score	Phoenix Sepsis Score		Phoenix-8 Score		pSOFA Score	
	Prompt assessment[a]	24-hour assessment[b]	Prompt assessment	24-hour assessment	Prompt assessment	24-hour assessment
In-hospital mortality	0.43 (0.32–0.54)	0.42 (0.31–0.55)	0.43 (0.32–0.55)	0.43 (0.31–0.55)	0.31 (0.22–0.41)	0.34 (0.24–0.46)
One-month mortality	0.72 (0.60–0.81)	0.72 (0.60–0.82)	0.72 (0.60–0.81)	0.72 (0.60–0.82)	0.67 (0.54–0.78)	0.67 (0.55–0.78)

AUROC

Score	Phoenix Sepsis Score		Phoenix-8 Score		pSOFA Score	
	Prompt assessment	24-hour assessment	Prompt assessment	24-hour assessment	Prompt assessment	24-hour assessment
In-hospital mortality	0.85 (0.80–0.90)	0.86 (0.81–0.91)	0.86 (0.81–0.90)	0.88 (0.83–0.92)	0.85 (0.81–0.89)	0.85 (0.80–0.89)
One-month mortality	0.87 (0.82–0.92)	0.89 (0.84–0.94)	0.89 (0.84–0.93)	0.90 (0.86–0.94)	0.88 (0.84–0.91)	0.87 (0.83–0.91)

Fig. 2. Comparison of the Phoenix Sepsis Score, the Phoenix-8 Score, and the pSOFA score based on their performance in predicting mortality outcomes in patients with suspected infection. Each score was calculated using 2 different methods, and the AUPRCs and AUROCs were calculated for each of the generated scores. AUPRCs and AUROCs are directly shown with 95% confidence intervals and are also presented with color hit maps: higher values are represented by darker colors. pSOFA = pediatric Sequential Organ Failure Assessment, AUPRC = area under the precision-recall curve, AUROC = area under the receiver operating characteristic curve.
[a]Representing scores calculated from sub-scores based on values closest to the PICU admission time.
[b]Representing scores calculated from worst sub-scores based on values collected within 24 hours after PICU admission.

https://doi.org/10.3346/jkms.2025.40.e106

The Phoenix Sepsis Score was originally introduced and validated using the worst sub-scores assessed in a window of 24 hours, similar to previously introduced organ dysfunction scores, such as pSOFA score.[10,11] While the intended interpretation would be comprehensive organ dysfunction status of the preceding 24 hours, extended application of the scores to more acute settings with shorter time windows may be beneficial, as pSOFA score was previously validated in a more limited emergency department setting.[6,12] Using data closest to admission times collected within a shorter time window was an attempt to simulate such scenarios for better representation of the acute state of patients presenting to PICUs. The resulting scores were not only comparable to those calculated from the worst values within 24 hours in terms of AUROC and AUPRC, but they also allocated fewer survivors to the highest scores, while maintaining overall sensitivity.

The limitations of this study include the retrospective design confined to a single PICU. Additionally, treating missing values and interventions as non-additive to scores could lead to an underestimation of organ dysfunction scores. However, it can be assumed that those variables were omitted in situations where their evaluation was unnecessary at the time. We only assessed performance of the scores on the first day of PICU admission, and further investigation on scores of subsequent days may be beneficial for better understanding their utility in monitoring critically ill children. Additionally, a significantly higher proportion of children with suspected infections compared to the original development study was observed. This discrepancy may be explained by the restriction of our study to PICU and the high prevalence of antibiotic use in Korea.[13,14] However, it also highlights an inherent limitation of the Phoenix criteria: the operational definition of suspected infection is influenced by clinicians' practice patterns. Another limitation of our study is the difference from the original study's protocol, specifically the exclusion of PICU stays shorter than 24 hours.[5] The excluded cases predominantly involve mortality events, with worst sepsis scores calculated immediately prior to death, reflecting terminal organ failure. This deviation may limit the comparability of our results.

The Phoenix sepsis criteria and the Phoenix Sepsis Score are reliable predictors of mortality outcomes in PICU patients with suspected infections. Prompt assessment at PICU admission exhibited comparable results to the 24-hour assessment. Further investigations to explore its applicability across diverse clinical settings are thus warranted.

Ethics statement

This study was approved by the Institutional Review Board (IRB) of Severance Hospital (Seoul, Korea; IRB No. 4-2024-0203). Informed consent was waived due to the retrospective design of the study.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Characteristics of PICU encounters with suspected infection

Supplementary Table 2
Performance metrics of the Phoenix Sepsis Score for mortality outcomes

https://doi.org/10.3346/jkms.2025.40.e106

Supplementary Fig. 1

Flow diagram.

REFERENCES

1. Schlapbach LJ, Watson RS, Sorce LR, Argent AC, Menon K, Hall MW, et al. International consensus criteria for pediatric sepsis and septic shock. *JAMA* 2024;331(8):665-74. PUBMED | CROSSREF
2. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *Lancet* 2020;395(10219):200-11. PUBMED | CROSSREF
3. Goldstein B, Giroir B, Randolph A; International Consensus Conference on Pediatric Sepsis. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 2005;6(1):2-8. PUBMED | CROSSREF
4. Menon K, Schlapbach LJ, Akech S, Argent A, Biban P, Carrol ED, et al. Criteria for pediatric sepsis-a systematic review and meta-analysis by the pediatric sepsis definition taskforce. *Crit Care Med* 2022;50(1):21-36. PUBMED | CROSSREF
5. Sanchez-Pinto LN, Bennett TD, DeWitt PE, Russell S, Rebull MN, Martin B, et al. Development and validation of the Phoenix criteria for pediatric sepsis and septic shock. *JAMA* 2024;331(8):675-86. PUBMED | CROSSREF
6. Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care* 2019;23(1):374. PUBMED | CROSSREF
7. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6(11):e012799. PUBMED | CROSSREF
8. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17(8):857-72. PUBMED | CROSSREF
9. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10(3):e0118432. PUBMED | CROSSREF
10. Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric Sequential Organ Failure Assessment score and evaluation of the Sepsis-3 definitions in critically ill children. *JAMA Pediatr* 2017;171(10):e172352. PUBMED | CROSSREF
11. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;315(8):801-10. PUBMED | CROSSREF
12. Balamuth F, Scott HF, Weiss SL, Webb M, Chamberlain JM, Bajaj L, et al. Validation of the pediatric Sequential Organ Failure Assessment score and evaluation of Third International Consensus Definitions for Sepsis and Septic Shock definitions in the pediatric emergency department. *JAMA Pediatr* 2022;176(7):672-8. PUBMED | CROSSREF
13. Choe YJ, Shin JY. Trends in the use of antibiotics among Korean children. *Korean J Pediatr* 2019;62(4):113-8. PUBMED | CROSSREF
14. Park SY, Moon SM, Kim B, Lee MJ, Park JY, Hwang S, et al. Appropriateness of antibiotic prescriptions during hospitalization and ambulatory care: a multicentre prevalence survey in Korea. *J Glob Antimicrob Resist* 2022;29:253-8. PUBMED | CROSSREF