

Preference-Based Assessments

Construction and Validation of a Risk Prediction Model for Prolonged Hospitalization of Very Premature Infants

Yang Yang, MD,* Huan Yang, BSN,* Hui Rong, MNS, Xian-wen Li, PhD, Rui Cheng, MD, Fei Shen, MNS

ABSTRACT

Objectives: This study aims to design a predictive model for extension of length of stay (LOS) in very preterm infants (VPIs), for risk management and assisted decision making in the early postnatal period.

Methods: VPIs in the development cohort were randomly divided into training and testing sets in a 7:3 ratio. A total of 5 machine learning algorithms were used to construct and evaluate the model. LOS extension was defined as exceeding the 75th percentile of total hospitalization days for different gestational age groups.

Results: This study included a total of 1044 VPIs in the development cohort, and 23.9% ($n = 250$) were classified as having LOS extension. Seven variables were screened and selected to construct the prediction model based on the best algorithm, logistic regression (LR). In the internal validation, compared with other algorithms, the LR algorithm achieved the highest area under the curve (AUC) of 0.773 (95% CI 0.717–0.830). The accuracy was 0.729, specificity was 0.782, recall was 0.566, and F1 score was 0.503. External validation of the LR model yielded an AUC value of 0.727 (95% CI 0.674–0.780). In terms of calibration curves, apart from the internal validation set showing a slight overestimation, both the training set and the external validation set demonstrated good consistency. Moreover, the decision curve analysis showed that the model has appropriate clinical applicability.

Conclusions: The predictive model could help healthcare professionals predict and address potential risks of LOS extension in VPIs.

Keywords: length of stay, machine learning, prediction model, risk factors, very preterm infants.

VALUE HEALTH. 2025; 28(10):1565–1573

Highlights

- Very premature infants often require long-term hospitalization, which imposes a heavy burden on parents, hospitals, and society. Currently, there remains a lack of consensus on how to effectively leverage the high-risk factors to develop predictive models for accurately identifying the extended hospital stay.
- This study constructed a prediction model for prolonged hospitalization. The model showed robust predictive ability and applicability in both internal and external validation, enabling early prediction of the risk of extended hospital stay within 72 hours after birth.
- This model may help medical staff make decisions and address in risk factors, leading to a possible reduction in the length of hospital stays, reducing parents' separation anxiety, and optimizing the resource management of neonatal intensive care units.

Introduction

Owing to their underdeveloped physiological state, very premature infants (VPIs) typically require prolonged hospitalization for comprehensive medical care. In China, approximately 200 000 VPIs (gestational age [GA] < 32 weeks) are born annually, with the majority receiving intensive treatment and care in neonatal intensive care units (NICUs). Over recent years, the mortality rate among VPIs has significantly decreased.^{1,2} Therefore, healthcare professionals have increasingly turned their attention to shortening the length of stay (LOS) and efficiently allocate medical resources in NICUs.^{3,4} Previous studies have demonstrated that LOS extension exposes infants to adverse environmental factors for prolonged periods, increasing the risk of hospital-acquired infections, escalating healthcare service burdens, and exacerbating maternal-infant separation anxiety.^{5–8} Thus, it is imperative to ensure effective treatment while simultaneously minimizing LOS in NICUs. One important challenge faced is predicting whether VPIs will experience LOS extension.

Previous studies have investigated the influencing factors of neonatal LOS. Seaton et al³ found that intrinsic factors of

newborns, particularly birth weight, GA, and gender, are the most significant predictors of LOS.³ Fu et al⁹ identified 58 potential risk factors affecting LOS in NICUs, encompassing maternal factors, prenatal treatment factors, neonatal diseases and treatments, neonatal clinical scores, and laboratory indicators. They further pinpointed 6 key risk factors: birth weight, GA, sepsis, necrotizing enterocolitis (NEC), bronchopulmonary dysplasia (BPD), and retinopathy of prematurity (ROP). Despite these findings, there remains a lack of consensus on how to effectively leverage these risk factors to develop predictive models for accurately identifying the extended hospital stay.

A recent study highlighted that most existing models for predicting neonatal LOS extension exhibit limited generalization ability and applicability in the hospital management.¹⁰ First, this limitation may stem from the fact that data sets used for model construction are predominantly derived from single centers with relatively small sample sizes, lacking external validation.^{11,12} Second, previous models for predicting neonatal LOS have primarily

*Huan Yang and Yang Yang are co-first authors.

relied on regression analysis methods, which may not be suitable for handling complicated variable types or may exclude potentially useful categorical variables.¹⁰ In contrast, machine learning (ML) approaches have been successfully applied to establish prediction models for LOS extension in other age groups, yielding promising results and benefits.^{11–13} To our knowledge, in recent years, only Lin et al¹⁴ and Frostig et al¹² have utilized ML algorithms to predict neonatal LOS extension. However, these studies still exhibit certain limitations, such as insufficient targeting of VPIs, lack of multicenter research design, and inadequate external validation.

Based on the aforementioned challenges, this study aims to develop an easy-to-use predictive model for LOS extension in VPIs using ML. External validation was performed to assess the accuracy and robustness of the model. Our objective is to provide a decision-support tool for both parents and medical staff after the early admission of neonates.

Methods

Data Source

The clinical data of all premature infants in this retrospective study were obtained from the Jiangsu provincial neonatal respiratory failure collaboration network (JSNRFNC) (ISRCTN registration number: 13622464). The data set was sourced from 30 tertiary NICUs across 13 cities in Jiangsu Province, which provide comprehensive neonatal services and can manage high-risk newborns. All participating centers adhere to standardized protocols for major clinical interventions, including neonatal resuscitation, tracheal intubation, mechanical ventilation, and administration of pulmonary surfactant.

Inclusion Criteria

This study cohort comprises premature infants with a GA of 24 to 31 weeks who were admitted to the NICUs of the JSNRFNC for the first time within 24 hours of birth, from January 1, 2019, to December 31, 2021.

Exclusion Criteria

Infants transferred more than 24 hours after birth were excluded because of incomplete medical records before transportation. Additionally, we excluded infants with severe congenital malformations, those transferred to hospitals outside the JSNRFNC, those requiring emergency surgery upon admission, those discharged against medical advice, or deceased infants.

Sample Size Estimation

The sample size was calculated using R software (4.2.1) based on the clinical prediction model guideline published by the British Medical Journal.^{15,16}

Development of the model: According to literature, the incidence of LOS extension in the VPI population is approximately 26%, assuming an incidence rate of 0.26 for each event per predictor parameter. According to Lin et al¹⁴, the R^2 in constructing a risk prediction model for LOS extension in the VPI population is 0.68. Assuming an acceptable difference of 0.05 between the model's apparent and adjusted R^2 , an intercept estimation error of 0.05, and 62 prediction factors, the minimum sample size required to develop a new model is 1037 cases (containing 270 events) using "pmsamplesize" package.

External validation of the model: (1) "pmsamplesize" method: expected area under the curve (AUC) value = 0.80, allowable difference (δ) = 0.05, significance level (α) = 0.05, power ($1-\beta$) = 0.80

(because $\beta = 0.20$). The calculation yields a minimum external validation sample size of 268 cases with 70 events. (2) Formula method: sample size = $(Z\alpha + Z\beta)^2 \times [p(1-p)]/\delta^2$, p : expected AUC value = 0.80, δ = 0.05, α = 0.05, β = 0.20. The sample size is around 246 cases with 64 events.^{17,18} (3) Rule of thumb: a commonly accepted rule of thumb suggests that a minimum of 100 positive events is required.^{17,18} Finally, the minimum sample size we adopted was at least 300 cases (including 100 events).

Predictive Variable

In the process of feature selection, we comprehensively evaluated the risk factors of the perinatal period based on both previous reports and existing variables in the JSNRFNC. The 62 variables we collected mainly include the following: (1) maternal data: home address (rural area or urban area), maternal age, pregnancy-induced hypertension, gestational diabetes mellitus, congenital heart disease, maternal severe anemic, chorioamnionitis, conception mode (nature conceived or assisted reproduction), number of fetuses, placental abruption, premature rupture of membranes, intrauterine distress, prenatal glucocorticoid use, placental abruption, delivery institution, delivery mode (vaginal delivery, selective C-section, and emergency C-section), amniotic fluid volume, amniotic fluid turbidity, etc; (2) neonatal data: GA, birth weight, small for gestational age (SGA), sex, 1-minute Apgar, 5-minute Apgar, resuscitation in delivery room (respiratory support by T-piece, self-inflation bag, and intubation, chest compression), score for neonatal acute physiology, perinatal extension, version II, age and weight at NICU transfer, admission diagnosis (respiratory distress syndrome, early-onset sepsis), respiratory support (noninvasive or invasive ventilation), nitric oxide inhalation, medications (Cedilanid, milrinone, adrenaline, analgesics, sedatives, dobutamine, dopamine, caffeine, surfactant, etc), clinical outcomes (air leak, patent ductus arteriosus, NEC, BPD, retinopathy of prematurity, metabolic bone disease, respiratory distress syndrome, meconium aspiration syndrome, late-onset sepsis, pneumonia, persistent pulmonary hypertension in newborns, hypoxic ischemic encephalopathy, intraventricular hemorrhage, central nervous system infection, periventricular leukomalacia), etc. All data were collected before birth or within 72 hours after birth.

Main Outcome

Because of variations in hospitalization duration among different GA groups, we define LOS extension as exceeding the 75th percentile of total hospitalization days for each respective GA group.¹⁴ In this study, the 75th percentile of total hospitalization days for premature infants with GAs of 25 to 31 weeks were as follows: 86 days for 25 weeks, 79 days for 26 weeks, 74 days for 27 weeks, 66 days for 28 weeks, 56 days for 29 weeks, 47 days for 30 weeks, and 38 days for 31 weeks.

Model Development, Validation, and Evaluation

In this study, 25 centers outside Nanjing city were classified into the development queue, and the other 5 centers in Nanjing city were classified into the external validation set. We used random stratified sampling to partition the development queue into training and testing sets in a 7:3 ratio. Five ML algorithms were primarily utilized—random forest (RF), light gradient boosting machine (LGBM), support vector machine (SVM), K-nearest neighbor (KNN), and logistic regression (LR)—to develop the predictive model. The trained model was subsequently validated and evaluated on the testing set. Additionally, an external validation set was used to assess the model's transportability and generalizability.

When evaluating the predictive performance of different models for LOS extension, the assessment focuses on 2 key aspects: discrimination and calibration. Discrimination is evaluated using the AUC of receiver operating characteristic (ROC) curve, whereas calibration is assessed using Brier scores and calibration curves (Bootstrap resampling, $n = 500$). A lower Brier score indicates better model calibration.¹⁹ Additionally, the net benefits of the model at various thresholds are evaluated through decision curve analysis (DCA). To facilitate clinical application, an online prediction platform has been developed. Sensitivity analysis was also conducted to verify the stability of the prediction model using 2 subgroups from the external validation set: (1) preterm infants with a GA < 30 weeks and (2) those with a birth weight < 1500 g.

Statistical Analysis

Statistical analysis was conducted using R version 4.2.1. Variables with more than 20% missing values were excluded from the analysis. For variables with less than 20% missing values, multiple imputation methods were used to handle the missing data.²⁰

Univariate analysis was used to describe the relationship between variables. Quantitative data following a normal distribution were presented as mean \pm standard deviation. Comparisons between 2 groups were performed using the *t* test. For skewed distributions, data were summarized by the median and interquartile range (IQR), and the Mann-Whitney U test was used for group comparisons. Qualitative data were analyzed using the Pearson Chi-square test or Fisher's exact test as appropriate.

The least absolute shrinkage and selection operator (LASSO) regression was used in the training set to screen and identify key predictive factors. This method enhances the prediction accuracy of ML models and mitigates overfitting.²¹ The selected predictive factors were incorporated into five ML models, and the optimal model was chosen based on AUCs, Brier scores, and calibration curves of different ML algorithms. In the optimal model, the DeLong test was used to assess significant differences in AUCs between different data sets.

Results

Baseline Characteristics

A total of 2142 VPIs were included in the JSNRFN. According to the inclusion and exclusion criteria, 1044 VPIs were included in the model development cohort, and 450 were used as an external validation cohort (see [Appendix Fig. 1 in Supplemental Materials](https://doi.org/10.1016/j.jval.2025.06.011) found at <https://doi.org/10.1016/j.jval.2025.06.011>). The median GA of VPIs in the development cohort was 30.43 weeks (IQR: 29.29, 31.14), with a median birth weight of 1,435 g (IQR: 1,200, 1,620). There were 609 males (58.3%), and the median LOS was 41 days (IQR: 32, 53). [Table 1](#) presents the baseline characteristics of the development and external validation cohorts. In the training and testing sets, baseline features were generally balanced and comparable. Compared with the development cohort, the external validation cohort had a significantly lower GA and birth weight.

We provided a detailed comparison between patients who experienced and did not experience LOS extension in the development cohort and external validation cohort in [Appendix Table 1 in Supplemental Materials](https://doi.org/10.1016/j.jval.2025.06.011) found at <https://doi.org/10.1016/j.jval.2025.06.011>. Among these, 23.9% ($n = 250$) of patients in the development cohort and 26.9% ($n = 121$) of patients in the external validation cohort were classified as having LOS extension, respectively.

Because of the smaller number of events in model development cohort, we used the “pwr” package to conduct a

supplemental statistical power analysis for the differences in the proportions between the LOS extension group and the non-LOS extension group. Effect size (Cohen's *h*) was set as 0.3, and significance level = 0.05. The calculated power was found to be 0.9852489.

Feature Selection and Parameter Adjustment

Before variable screening, we conducted an analysis using a correlation matrix and found that the variables to be included had no obvious correlation (see [Appendix Fig. 2 in Supplemental Materials](https://doi.org/10.1016/j.jval.2025.06.011) found at <https://doi.org/10.1016/j.jval.2025.06.011>). Then, LASSO regression was used for feature selection, and 10-fold cross-validation was used to determine the optimal lambda (λ) value, as shown in [Appendix Table 2](#) and [Appendix Figures 3 and 4 in Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2025.06.011>. When the contraction parameter lambda.1se was set to 0.03776389, 9 of the 62 predictive indicators were selected by LASSO, including age at NICU transfer, SGA, weight at NICU admission, fetal distress, 1-minute Apgar score, pregnancy-induced hypertension, initiation of invasive ventilation within 72 hours after birth, use of surfactant ≥ 2 times, and administration of Cedilanid.

Development and Comparison of Predictive Models

RF, LGBM, SVM, KNN, and LR algorithms were trained on the training set of the development cohort and internally validated using the testing set. To predict LOS extension, ROC curves were generated for each of the 5 models.

In internal validation, the AUC of the LR model was the highest at 0.773 (95% CI 0.717-0.830). The AUCs for the RF, LGBM, SVM, and KNN models were 0.760 (95% CI 0.703-0.818), 0.740 (95% CI 0.682-0.799), 0.734 (95% CI 0.671-0.797), and 0.688 (95% CI 0.623-0.754), respectively. To comprehensively evaluate the models, we calculated accuracy, specificity, recall, F1 score, and Brier score, as presented in [Table 2](#). Because of its highest AUC and lowest Brier score, the LR model was ultimately selected to predict LOS extension in VPIs. Its performance metrics were as follows: accuracy = 0.729, specificity = 0.782, recall = 0.566, F1 score = 0.503, and Brier score = 0.152. The detailed ROC, calibration curves, and DCA for the 5 ML models are shown in [Appendix Figure 5 in Supplemental Materials](https://doi.org/10.1016/j.jval.2025.06.011) found at <https://doi.org/10.1016/j.jval.2025.06.011>.

Internal and External Validation

Based on the LR model using the backward method, 7 of the 9 variables first selected by LASSO were further screened: age at NICU transfer, SGA, weight at NICU admission, fetal distress, pregnancy-induced hypertension, initiation of invasive ventilation within 72 hours after birth, and use of Cedilanid, as shown in [Table 3](#). External validation of the LR model yielded an AUC value of 0.727 (95% CI 0.674-0.780), as presented in [Figure 1](#). The DeLong test indicated no significant difference in AUCs between the internal validation set and the training set, and between the external set and the training set ($P = .3396$). In terms of calibration curves, apart from the internal validation set showing a certain degree of overestimation for the extremely low-risk subgroup, both the training set and the external validation set demonstrated good consistency between the predicted probabilities and the actual probabilities. Across training, internal validation, and external validation sets, calibration intercepts were measured at 0.000, -0.011, and 0.000 with corresponding slopes of 1.000, 1.001, and 1.000 (see [Appendix Table 3 in Supplemental Materials](https://doi.org/10.1016/j.jval.2025.06.011) found at <https://doi.org/10.1016/j.jval.2025.06.011>). Meanwhile, both the Brier scores and the mean absolute error showed that the

Table 1. Univariate analysis between the test set, training set, and external validation set.

Variables	Development queue (<i>n</i> = 1044)				External validation queue (<i>n</i> = 450)		<i>P</i> value [†]
	Total (<i>n</i> = 1044)	Training set (<i>n</i> = 730)	Test set (<i>n</i> = 314)	<i>P</i>	External validation set (<i>n</i> = 450)	<i>P</i> *	
Inherent factors							
Male, <i>n</i> (%)	609 (58)	430 (59)	179 (57)	.616	256 (57)	.535	.644
Birth weight, Median (Q1,Q3), g	1435 (1200, 1620)	1440 (1220, 1620)	1430 (1200, 1610)	.587	1313.85 ± 309.93	<.001	<.001
Gestational age, Median (Q1,Q3), weeks	30.43 (29.29, 31.14)	30.29 (29.29, 31.14)	30.43 (29.14, 31.29)	.839	30 (28.71, 31)	<.001	<.001
SGA, <i>n</i> (%)	38 (4)	28 (4)	10 (3)	.738	24 (5)	.284	.172
In vitro fertilization, <i>n</i> (%)	156 (15)	116 (16)	40 (13)	.224	86 (19)	.178	.054
Multiple gestation, <i>n</i> (%)	244 (23)	171 (23)	73 (23)	1	118 (26)	.31	.265
Antenatal treatment and maternal factors							
Prenatal glucocorticoids, <i>n</i> (%)	855 (82)	597 (82)	258 (82)	.952	389 (86)	.044	.037
Full course of prenatal glucocorticoids, <i>n</i> (%)	451 (43)	314 (43)	137 (44)	.907	264 (59)	<.001	<.001
Fetal distress, <i>n</i> (%)	97 (9)	70 (10)	27 (9)	.697	17 (4)	<.001	<.001
Emergency cesarean section, <i>n</i> (%)	217 (21)	146 (20)	71 (23)	.384	155 (34)	<.001	<.001
Spontaneous delivery, <i>n</i> (%)	455 (44)	326 (45)	129 (41)	.317	186 (41)	.29	.454
Bloody amniotic fluid, <i>n</i> (%)	55 (5)	40 (5)	15 (5)	.753	20 (4)	.516	.589
Normal amniotic fluid volume, <i>n</i> (%)	956 (92)	669 (92)	287 (91)	.994	424 (94)	.126	.096
Maternal age, Median (Q1,Q3), y	30 (27, 33)	30 (27, 33)	30 (27, 33.75)	.467	31 (28, 34)	.074	.03
Maternal diabetes, <i>n</i> (%)	173 (17)	124 (17)	49 (16)	.646	66 (15)	.331	.399
Placental abruption, <i>n</i> (%)	56 (5)	45 (6)	11 (4)	.11	16 (4)	.067	.172
Premature rupture of membranes >24 h, <i>n</i> (%)	192 (18)	130 (18)	62 (20)	.513	43 (10)	<.001	<.001
Pregnancy-induced hypertension, <i>n</i> (%)	152 (15)	105 (14)	47 (15)	.881	75 (17)	.329	.336
Mother with chronic hypertension, <i>n</i> (%)	24 (2)	21 (3)	3 (1)	.094	8 (2)	.322	.657
Chorioamnionitis, <i>n</i> (%)	95 (9)	64 (9)	31 (10)	.651	18 (4)	.003	<.001
Maternal severe anemia, <i>n</i> (%)	13 (1)	9 (1)	4 (1)	1	2 (0)	.222	.256
Conditions of the baby							
1min Apgar score, Median (Q1,Q3)	8 (6, 8)	8 (6, 8)	8 (7, 9)	.12	8 (7, 9)	<.001	<.001
5min Apgar score, Median (Q1,Q3)	8 (8, 9)	8 (8, 9)	8 (8, 9)	.382	9 (8, 10)	<.001	<.001
Admission age to NICU, Median (Q1,Q3), hours	0.3 (0.16, 0.7)	0.3 (0.16, 0.68)	0.31 (0.16, 0.78)	.519	1 (0.1, 1)	<.001	<.001
Weight when transferred to NICU, Median (Q1,Q3), g	1430 (1200, 1610)	1435 (1220, 1610)	1430 (1200, 1607.5)	.565	1310.58 ± 309.36	<.001	<.001
RDS, <i>n</i> (%)	713 (68)	498 (68)	215 (68)	.994	340 (76)	.008	.006
Sepsis, <i>n</i> (%)	273 (26)	189 (26)	84 (27)	.831	171 (38)	<.001	<.001
SNAPPE II, Median (Q1,Q3)	14 (5, 20)	14 (5, 21)	14 (7, 20)	.987	8 (5, 12)	<.001	<.001
Treatment of the baby							
Resuscitation. self-inflating bag, <i>n</i> (%)	280 (27)	202 (28)	78 (25)	.384	72 (16)	<.001	<.001

continued on next page

Table 1. Continued

Variables	Development queue (n = 1044)				External validation queue (n = 450)		P value [†]
	Total (n = 1044)	Training set (n = 730)	Test set (n = 314)	P	External validation set (n = 450)	P*	
Resuscitation. T-piece, n (%)	34 (3)	22 (3)	12 (4)	.628	131 (29)	<.001	<.001
Resuscitation. Intubation, n (%)	96 (9)	68 (9)	28 (9)	.93	39 (9)	.785	.819
Resuscitation. Compression, n (%)	48 (5)	32 (4)	16 (5)	.732	6 (1)	.007	.003
Resuscitation. Surfactant, n (%)	8 (1)	6 (1)	2 (1)	1	2 (0)	.718	.732
Noninvasive MV within 24 h after birth, n (%)	801 (77)	560 (77)	241 (77)	1	338 (75)	.578	.545
Noninvasive MV within 72 h after birth, n (%)	866 (83)	608 (83)	258 (82)	.725	369 (82)	.624	.711
Noninvasive MV failure within 24 h after birth, n (%)	47 (5)	28 (4)	19 (6)	.155	49 (11)	<.001	<.001
Noninvasive MV failure within 72 h after birth, n (%)	67 (6)	40 (5)	27 (9)	.08	59 (13)	<.001	<.001
Invasive MV within 24 h after birth, n (%)	282 (27)	189 (26)	93 (30)	.243	166 (37)	<.001	<.001
Invasive MV within 72 h after birth, n (%)	303 (29)	202 (28)	101 (32)	.164	176 (39)	<.001	<.001
Used iNO within 24 h after birth, n (%)	5 (0)	1 (0)	4 (1)	.031	3 (1)	.158	.704
Used iNO within 72 h after birth, n (%)	8 (1)	4 (1)	4 (1)	.251	3 (1)	1	1
Used PS, n (%)	556 (53)	394 (54)	162 (52)	.523	272 (60)	.034	.012
Used PS \geq 2 doses, n (%)	90 (9)	63 (9)	27 (9)	1	38 (8)	.997	.991
Caffeine, n (%)	710 (68)	506 (69)	204 (65)	.191	398 (88)	<.001	<.001
Dopamine, n (%)	216 (21)	147 (20)	69 (22)	.556	87 (19)	.794	.597
Dobutamine, n (%)	135 (13)	91 (12)	44 (14)	.56	80 (18)	.015	.018
Cedilanid, n (%)	19 (2)	15 (2)	4 (1)	.54	17 (4)	.113	.038
Milrinone, n (%)	12 (1)	9 (1)	3 (1)	1	0 (0)	.016	.023
Adrenaline, n (%)	12 (1)	8 (1)	4 (1)	.76	2 (0)	.334	.251
Used \geq 2 vasoactive drugs, n (%)	235 (23)	162 (22)	73 (23)	.769	108 (24)	.518	.575
Sedatives, n (%)	60 (6)	45 (6)	15 (5)	.46	8 (2)	<.001	.001
Analgesic drugs, n (%)	4 (0)	3 (0)	1 (0)	1	1 (0)	1	1
Organizational factors							
Class III hospital delivery, n (%)	982 (94)	684 (94)	298 (95)	.54	437 (97)	.013	.019
Living in urban area, n (%)	718 (69)	493 (68)	225 (72)	.213	345 (77)	<.001	.002
Gestational age at discharge, Median (Q1,Q3), weeks	36.14 (35.29, 37.29)	36.14 (35.14, 37.29)	36.14 (35.29, 37.14)	.785	36.14 (35, 37.43)	.905	.848
Weight at discharge, Median (Q1,Q3), g	2210 (2070, 2440)	2210 (2070, 2428.75)	2220 (2080, 2463.75)	.452	2110 (1920, 2320)	<.001	<.001
Length of stay, Median (Q1,Q3), day	41 (32, 53)	41 (32, 52.75)	41.5 (32, 56)	.798	45 (33, 59)	.005	.004

iNO indicates inhaled NO; MV, mechanical ventilation; NICU, neonatal intensive care unit; PS, pulmonary surfactant; RDS, respiratory distress syndrome; SGA, small for gestational age; SNAPPE II, score for neonatal acute physiology with perinatal extension-II.

*External validation set versus train set.

[†]Development queue versus external validation queue.

Table 2. Detailed parameters for constructing models using different algorithms.

ML Type	AUC	Accuracy	Specificity	Recall	F1 score	Brier score
LR	0.773	0.729	0.782	0.566	0.503	0.152
RF	0.760	0.694	0.719	0.618	0.495	0.160
LGBM	0.740	0.637	0.584	0.803	0.517	0.163
SVM	0.734	0.713	0.777	0.513	0.464	0.172
KNN	0.688	0.72	0.849	0.316	0.353	0.206

AUC indicates area under curve; KNN, K-nearest neighbor; LGBM, light gradient boosting machine; LR, logistic regression; ML, machine learning; RF, random forest; SVM, support vector machine.

model probability prediction has high accuracy. (Brier score for the training set: 0.145, testing set: 0.152, and external validation set: 0.175; Mean absolute error for the training set: 0.021, testing set: 0.032, and external validation set: 0.015) (see [Appendix Fig. 6](#) in [Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2025.06.011>). DCA for these sets showed that the model has appropriate clinical applicability (see [Appendix Fig. 7](#) in [Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2025.06.011>).

Web Nomogram

Based on the results of LASSO regression and LR, a prediction model was constructed using the above 7 predictive factors. The model is presented as a web-based nomogram, available at <https://shenfei.shinyapps.io/dynnomapp/>. For example, if a VPI is admitted to the NICU within 12 hours of birth, the mother has pregnancy-induced hypertension, there is a history of fetal distress, the infant's weight at NICU admission is 980 g, the infant is SGA, and invasive mechanical ventilation (IMV) or Cedilanid treatment has not been initiated within 72 hours after birth, the predicted probability of LOS extension is 98.3% (95% CI 91.2%-99.7%).

Sensitivity Analysis

We conducted sensitivity analysis using 2 subgroup sets from the external validation cohort: (1) preterm infants with a GA < 30 weeks, and (2) preterm infants with a birth weight < 1500 g. The results demonstrated that both subgroups exhibited good AUCs and calibration accuracy. Specifically, the Brier scores for preterm infants with a birth weight < 1500 g and those with a GA < 30 weeks were 0.188 and 0.189, respectively (see [Appendix Fig. 8](#) in

[Supplemental Materials](#) found at <https://doi.org/10.1016/j.jval.2025.06.011>).

Discussion

In this multicenter study, we constructed a predictive model using 5 ML algorithms to identify the risk of LOS extension in the NICU within 3 days of birth for VPIs. Previous studies have treated LOS as a continuous variable rather than a categorical variable, which allows for direct prediction of specific LOS values.²²⁻²⁴ However, a recent multicenter study has confirmed significant regional differences in LOS,²⁵ making it challenging to consider LOS as a continuous variable in prediction models. Converting LOS into an ordered multiclass variable using the interquartile range method may be more reasonable for different regions,^{14,26} as LOS baselines vary across different regions and GA groups. This approach enables the use of a consistent definition or standard for statistical analysis across diverse settings. Lin et al¹⁴ expressed a similar viewpoint in their study, noting that converting LOS into categorical variables allows ML algorithms to establish more accurate models for predicting LOS extension. It is worth noting that the method of treating LOS as a categorical variable for prediction has also been applied and well validated in other populations.¹³

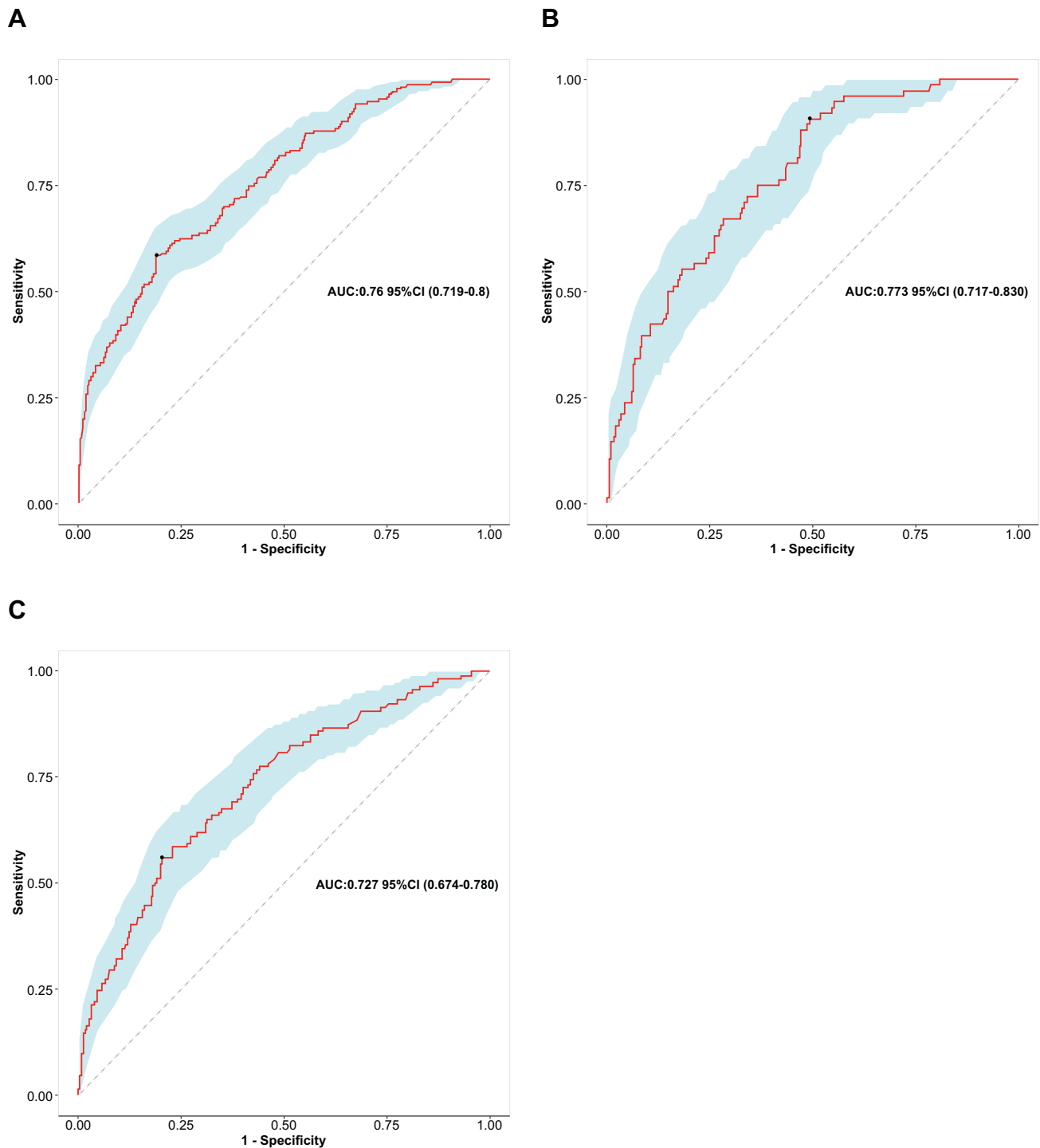
Our research results demonstrated that the LR model exhibited superior predictive performance compared with the other 4 ML models, with an AUC of 0.773 (95% CI 0.717-0.830), making it our optimal model. In external validation, the LR model maintained robustness and good predictive ability, achieving an AUC of 0.727 (95% CI 0.674-0.780). Regarding the calibration curve, whereas the internal validation set indicated some degree of overfitting, the independent external validation set, along with the sensitivity analysis, further help support the robustness of the model. The

Table 3. The selected variables through logistic regression analysis.

Characteristics	B	SE	OR(95%CI)	Z	P
(Intercept)	0.466	0.530	1.593 (0.565-4.535)	0.878	.380
Age at transferred to the NICU	0.144	0.055	1.155 (1.042-1.296)	2.614	.009
SGA	1.550	0.559	4.709 (1.665-15.51)	2.772	.006
Weight at transferred to the NICU	-0.002	<0.001	0.998 (0.997-0.999)	-4.626	<.001
Fetal distress	1.061	0.291	2.889 (1.623-5.106)	3.641	<.001
Maternal gestational hypertension	0.897	0.255	2.451 (1.479-4.026)	3.519	<.001
Started invasive ventilation within 72 hours after birth	0.608	0.207	1.837 (1.219-2.753)	2.933	.003
Use of Cedilanid	1.471	0.674	4.355 (1.207-18.02)	2.182	.029

CI indicates confidence interval; NICU, neonatal intensive care unit; OR, odds ratio; SE, standard error; SGA, small for gestational age.

Figure 1. ROC of the prediction risk model. (A) Training set. (B) Internal validation set. (C) External validation set.



AUC indicates area under the curve.

Brier score indicated that the LR model's predicted probabilities were consistent with actual probabilities, whereas DCA confirms its clinical applicability. Compared with previous studies, our model surpasses the one reported by Lin et al,¹⁴ which achieved an AUC of 0.724. This improvement may be attributed to the larger number of candidate variables included in our analysis. For feature selection, we adopted a combination of knowledge-driven and

data-driven approaches, resulting in a more comprehensive and efficient modeling foundation. Given the importance of early prediction,⁹ we excluded factors occurring 72 hours after birth, such as BPD, NEC, ROP, and other complications. Frostig et al¹² also confirmed that early prediction is more valuable than later prediction. However, recognizing that incorporating more data from later stages of NICU hospitalization could further enhance

predictive accuracy, our future work will consider including additional late-stage factors to develop LOS extension prediction models for different time periods of NICU hospitalization.

Our final model included 7 important predictive factors: age at NICU transfer, SGA, weight at NICU admission, fetal distress, pregnancy-induced hypertension, initiation of IMV within 72 hours postpartum, and use of Cedilanid. Among these factors, SGA, fetal distress, and pregnancy-induced hypertension are well-established contributors to neonatal morbidity and mortality and have been key predictors of LOS extension in previous studies.²⁷⁻²⁹ In our multivariate analysis, the risk of LOS extension was approximately 4.7 times higher for SGA patients compared with non-SGA patients, whereas fetal distress and pregnancy-induced hypertension increased the risks by approximately 2.9 and 2.5 times, respectively. Our findings align with a recent national survey,³⁰ which also identified SGA and maternal hypertension as significant factors associated with increased LOS. Birth weight is widely recognized as one of the most critical perinatal data for predicting LOS in the early postnatal period.³ In our study, weight at NICU admission replaced birth weight in the early prediction model. Given the complex environment VPIs face immediately after delivery, including the need for immediate warmth and resuscitation, guidelines recommend transferring babies to the NICU to stabilize body temperature before measuring weight.³¹ This approach may result in more accurate weight measurements compared with initial birth weight estimates, which can be imprecise. IMV is a common respiratory support used for critically ill newborns, indicating that VPIs requiring IMV often have more severe developmental issues and disease severity. Hintz et al²⁶ found that the need for respiratory support, including IMV within 24 hours after birth, is significantly associated with longer LOS in premature infants. These key variables are often selected and applied to similar predictive models.^{3,26-29} Among them, pregnancy-induced hypertension reflects the health factors of the pregnant mother. Age, weight, and SGA highlight the inherent factors of newborns. In addition, fetal distress and early initiation of IMV indicate the severity of the neonate's condition. Therefore, these predictive variables and the model can be widely used across different regions of China.

Interestingly, we also identified 2 previously underreported independent risk factors. Although the overall usage rate in our cohort is not high, the administration of Cedilanid within 72 hours after birth effectively predicted the risk of LOS extension in VPIs. Cedilanid, a digitalis-positive inotropic drug commonly used in patients with cardiac dysfunction, can indirectly reflect the impact of existing cardiac issues on outcome prediction.³² It is noteworthy that there is currently no unified guideline or standardized protocol regarding the optimal timing for Cedilanid administration. Consequently, the predictive performance of this variable may vary across different regions in China. Additionally, we found that delayed NICU admission is associated with a higher risk of LOS extension. Because of immature organ development and disease complications, most VPIs require immediate transfer to the NICU for treatment and intensive care after birth. Delayed NICU admission may be related to prolonged postnatal resuscitation or the inability of the delivery hospital to meet treatment needs, necessitating transfer to a higher-level NICU.³³ This study further confirms that pregnant women with high-risk factors in China should be prioritized for intrauterine transfer to medical institutions equipped to handle both high-risk pregnancies and VPI care.

Limitations

Our model has some limitations that warrant acknowledgment. We recognize that the prediction of LOS evolves with the

clinical progression of patients,³⁴ necessitating the dynamic incorporation of additional factors after 72 hours to update risk predictions at different stages. This is an important focus for our future work. Additionally, because of the retrospective study design and the inherent limitations of existing databases, certain social demographic factors and family involvement in caregiving, such as household income, parental psychological status, and kangaroo care, were not included in our analysis. As an important social event, COVID-19 led China to implement stricter control measures from 2020 to 2023. To some extent, this affects the length of hospital stay for newborns such as by limiting kangaroo care. Furthermore, Class imbalance is indeed an issue to consider. A majority to minority ratio of 3:1 (ie, minority < 25%) may produce a slight class imbalance. In this study, 23.9% ($n = 250$) of patients in the development cohort and 26.9% ($n = 121$) of patients in the external validation cohort were classified as having LOS extension, respectively. Although this impact of class imbalance may be small in our study, readers should remain cautious about the potential instability of results caused by this issue. Moreover, it should be noted that there were relatively few positive events in the model construction of this study. Although we used the LASSO method and cross-validation for variable screening and conducted internal and external validations, the impact of these methods still needs to be noted. Last but not least, as a regional study, the model validated externally needs further verification on a wider geographical scale. The geographic external validation in this study can to some extent test the regional adaptability of the model, but its essence is "limited external validation" and cannot fully simulate the complexity of the real world. Consequently, a prospective multicenter study across China could be conducted in the future.

Conclusions

This model demonstrated robust predictive ability and applicability in both internal and external validation, helping in the early prediction of the risk of LOS extension within 72 hours after VPIs are born, benefiting parents and clinical staff. This may help medical staff make decisions and address risk factors, thereby possibly shortening hospital stays, reducing parents' anxiety, and optimizing NICU resource management. Further research is still needed to validate and potentially refine our findings and conclusions.

Author Disclosures

Author disclosure forms can be accessed below in the [Supplemental Material](#) section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.jval.2025.06.011>.

Article and Author Information

Accepted for Publication: June 24, 2025

Published Online: July 24, 2025

doi: <https://doi.org/10.1016/j.jval.2025.06.011>

Author Affiliations: Department of Neonatology, Children's Hospital of Nanjing Medical University, Nanjing, Jiangsu, PR China (Y. Yang, H. Yang, Rong, Cheng, Shen); College of Nursing, Nanjing Medical University, Nanjing, Jiangsu, PR China (Li, Shen).

Correspondence: Rui Cheng, MD, Department of Neonatology, Children's Hospital of Nanjing Medical University, Nanjing, Jiangsu, PR China, 210008. Email: yy860508@163.com or chengrui350@163.com

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: The authors received no financial support for this research.

Acknowledgment: The authors thank the nurses for collecting and checking the data.

Ethical Approval: The ethics committee of the Children's Hospital of Nanjing Medical University approved this study (Number: 202401020-1). The institutional review board committee exempted the study from informed consent because of its retrospective nature. All data were totally anonymized before further statistical analysis. All the procedures were followed according to the Declaration of Helsinki.

Data Availability: The data set used during this study is available from the corresponding author upon reasonable request.

REFERENCES

- Cao Y, Jiang S, Sun J, et al. Assessment of Neonatal Intensive Care Unit practices, morbidity, and mortality among very preterm infants in China. *JAMA Netw Open*. 2021;4(8):e2118904.
- Jiang HT, Chen C. Survival status and trends of premature infants of different gestational ages. *Chin J Perinat Med*. 2024;27(10):865–870.
- Seaton SE, Barker L, Jenkins D, Draper ES, Abrams KR, Manktelow BN. What factors predict length of stay in a neonatal unit: a systematic review. *BMJ Open*. 2016;6(10):e010466.
- Lui K, Lee SK, Kusuda S, et al. Trends in outcomes for neonates born very preterm and very low birth weight in 11 high-income countries. *J Pediatr*. 2019;215:32–40.e14.
- Rivkees SA, Mayes L, Jacobs H, Gross I. Rest-activity patterns of premature infants are regulated by cycled lighting. *Pediatrics*. 2004;113(4):833–839.
- Caporali C, Pisoni C, Gasparini L, et al. A global perspective on parental stress in the neonatal intensive care unit: a meta-analytic study. *J Perinatol*. 2020;40(12):1739–1752.
- Rolnitsky A, Unger SL, Urbach DR, Bell CM. Cost of neonatal intensive care for extremely preterm infants in Canada. *Transl Pediatr*. 2021;10(6):1630–1636.
- Shields LB, Davydov Y, Glyder A, Weymouth C, Udwin M, Eakins M. Impact of technology on Neonatal Intensive Care Unit admissions and length of stay: a retrospective study. *Cureus*. 2023;15(6):e40813.
- Fu M, Song W, Yu G, Yu Y, Yang Q. Risk factors for length of NICU stay of newborns: a systematic review. *Front Pediatr*. 2023;11:1121406.
- Medeiros NB, Fogliatto FS, Rocha MK, Tortorella GL. Forecasting the length-of-stay of pediatric patients in hospitals: a scoping review. *BMC Health Serv Res*. 2021;21(1):938.
- Gokhale S, Taylor D, Gill J, et al. Hospital length of stay prediction tools for all hospital admissions and general medicine populations: systematic review and meta-analysis. *Front Med (Lausanne)*. 2023;10:1192969.
- Frostig T, Benjamini Y, Kehat O, et al. Developing a length of stay prediction model for newborns, achieving better accuracy with greater usability. *Int J Med Inform*. 2023;180:105267.
- Gokhale S, Taylor D, Gill J, et al. Hospital length of stay prediction for general surgery and total knee arthroplasty admissions: systematic review and meta-analysis of published prediction models. *Digit Health*. 2023;9:20552076231177497.
- Lin WT, Wu TY, Chen YJ, Chang YS, Lin CH, Lin YJ. Predicting in-hospital length of stay for very-low-birth-weight preterm infants using machine learning techniques. *J Formos Med Assoc*. 2022;121(6):1141–1148.
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
- Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
- Pavlou M, Qu C, Omar RZ, et al. Estimation of required sample size for external validation of risk models for binary outcomes. *Stat Methods Med Res*. 2021;30(10):2187–2206.
- Steyerberg EW. *Clinical Prediction Models: a Practical Approach to Development, Validation, and Updating*. 2nd ed. Cham, Switzerland: Springer; 2019.
- Hostetler ZS, Hsu FC, Yoganandan N, et al. An improved method for developing injury risk curves using the brier metric score. *Ann Biomed Eng*. 2021;49(11):3091–3098.
- Pan S, Chen S. Empirical comparison of imputation methods for multivariate missing data in public health. *Int J Environ Res Public Health*. 2023;20(2):1524.
- Liu X, Xie Z, Zhang Y, et al. Machine learning for predicting in-hospital mortality in elderly patients with heart failure combined with hypertension: a multicenter retrospective study. *Cardiovasc Diabetol*. 2024;23(1):407.
- Bender GJ, Koestler D, Ombao H, et al. Neonatal intensive care unit: predictive models for length of stay. *J Perinatol*. 2013;33(2):147–153.
- Hinchliffe SR, Seaton SE, Lambert PC, Draper ES, Field DJ, Manktelow BN. Modelling time to death or discharge in neonatal care: an application of competing risks. *Paediatr Perinat Epidemiol*. 2013;27(4):426–433.
- Lee HC, Bennett MV, Schulman J, Gould JB, Profit J. Estimating length of stay by patient type in the neonatal intensive care unit. *Am J Perinatol*. 2016;33(8):751–757.
- Seaton SE, Draper ES, Adams M, et al. Variations in neonatal length of stay of babies born extremely preterm: an international comparison between iNeo networks. *J Pediatr*. 2021;233:26–32.e6.
- Hintz SR, Bann CM, Ambalavanan N, et al. Predicting time to hospital discharge for extremely preterm infants. *Pediatrics*. 2010;125(1):e146–e154.
- Minor KC, Bianco K, Sie L, Druzin ML, Lee HC, Leonard SA. Severity of small-for-gestational-age and morbidity and mortality among very preterm neonates. *J Perinatol*. 2023;43(4):437–444.
- Cao C, Cai W, Niu X, et al. Prehypertension during pregnancy and risk of small for gestational age: a systematic review and meta-analysis. *J Matern Fetal Neonatal Med*. 2020;33(8):1447–1454.
- Lee HC, Bennett MV, Schulman J, Gould JB. Accounting for variation in length of NICU stay for extremely low birth weight infants. *J Perinatol*. 2013;33(11):872–876.
- Zhang M, Wang YC, Feng JX, et al. Variations in length of stay among survived very preterm infants admitted to Chinese neonatal intensive care units. *World J Pediatr*. 2022;18(2):126–134.
- Subspecialty Group of Neonatology, the Society of Pediatrics, Chinese Medical Association, Editorial Board, Chinese Journal of Pediatrics. Neonatology group of the Chinese Medical Association science branch, editorial committee of the Chinese journal of pediatrics. Clinical practice guidelines for resuscitation of premature infants with birth age < 32 weeks (2022). *Chin J Pediatr*. 2023;61(1):6–15.
- Jiang YJ, Zhang J, Wang JY, Liu W. Risk factors analysis and column chart prediction model construction for postoperative atrial fibrillation in critically ill patients undergoing lung surgery. Article in Chinese. *Chin J Thorac Cardiovasc Surg*. 2023;39(6):352–359.
- Neonatologist Association, Chinese Medical Doctor Association. Consensus guidelines for neonatal transfer (2017). Article in Chinese. *J Dev Med (Electron Version)*. 2017;5(4):193–197.
- Mehretie Y, Amare AT, Getnet GB, Mekonnen BA. Length of hospital stay and factors associated with very-low-birth-weight preterm neonates surviving to discharge: a cross-sectional study, 2022. *BMC Pediatr*. 2024;24(1):80.