

Original Research

Multiparametric Neonatal Lung Ultrasound Within Six Hours After Birth Improves Prediction of Respiratory Support Escalation in Preterm Infants: A Prospective Cohort Study

Pin Wang^{1,†}, Lihong Duan^{2,†}, Yanju Wang¹, Lisi Zhang¹, Yangyang Song¹, Yu Chen^{1,*} 

¹Department of Ultrasound, Shijiazhuang No.4 Hospital, 050000 Shijiazhuang, Hebei, China

²Department of Obstetrics, Shijiazhuang No.4 Hospital, 050000 Shijiazhuang, Hebei, China

*Correspondence: yu-chen@ldy.edu.rs (Yu Chen)

†These authors contributed equally.

Academic Editor: Michael H. Dahan

Submitted: 27 December 2025 Revised: 14 February 2026 Accepted: 26 February 2026 Published: 2 June 2026

Abstract

Background: Early prediction of respiratory support escalation in preterm infants remains challenging. Although the lung ultrasound score (LUS) assesses lung aeration, it does not capture key elements of respiratory muscle function or pulmonary circulatory adaptation. We evaluated whether a multiparametric ultrasound protocol, performed within ≤ 6 hours after birth, integrating LUS, diaphragmatic ultrasound, and pulmonary artery Doppler assessment, improves the prediction of respiratory support escalation within 72 hours. **Methods:** In this single-center prospective cohort study, preterm infants (26+0 to 33+6 weeks) underwent a standardized bedside ultrasound examination within ≤ 6 hours after birth. The protocol included LUS, assessment of diaphragmatic excursion and thickening fraction, as well as measurement of the pulmonary artery acceleration time-to-ejection time (AT/ET) ratio. Three predictive models were developed and compared: Model A (clinical variables only), Model B (clinical + LUS), and Model C (clinical variables + LUS + diaphragm ultrasound + AT/ET). The primary outcome was escalation of respiratory support within 72 hours after birth. Model performance was evaluated by discrimination (area under the curve [AUC], DeLong test), calibration (calibration slope, Brier score), and clinical utility (decision curve analysis). Measurement reproducibility was assessed using intraclass correlation coefficients (ICCs). **Results:** Among the 203 infants included in the analysis, 82 (40.4%) met the primary outcome. The complete multiparametric model (Model C) demonstrated significantly superior discriminative performance (AUC: 0.88, 95% confidence interval [CI]: 0.83–0.93) compared with Model A (AUC: 0.74) and Model B (AUC: 0.82) (both DeLong $p < 0.05$). Model C demonstrated excellent calibration (calibration slope: 0.99, Brier score: 0.13) and provided a higher net benefit than the simpler models across decision thresholds ranging from 0.22 to 0.50. Key ultrasound measurements showed excellent reproducibility (all ICCs ≥ 0.87). In Model C, lower gestational age, incomplete antenatal corticosteroid treatment, higher LUS, and lower diaphragm excursion, diaphragmatic thickening fraction, and AT/ET ratio were independently associated with an increased risk of respiratory support escalation. **Conclusions:** A multiparametric ultrasound assessment performed within 6 hours after birth significantly improves the early prediction of respiratory support escalation in preterm infants, offering high reproducibility and greater clinical net benefit than simpler predictive models.

Keywords: preterm infants; lung ultrasound; diaphragm ultrasound; pulmonary blood flow Doppler; predictive model; respiratory support

1. Introduction

The transition to extrauterine life presents a profound respiratory challenge for preterm infants. Successful adaptation during the first hours after birth is a critical determinant of short-term morbidity, setting the trajectory for the early course in the neonatal intensive care unit (NICU) [1]. This challenge is driven by multiple factors related to pulmonary immaturity, including underdeveloped alveolar architecture, surfactant deficiency, delayed clearance of fetal lung fluid, and a highly compliant chest wall [2]. The resulting pathophysiology, characterized by inadequate lung recruitment, ventilation-perfusion mismatch, and increased work of breathing, manifests clinically as persistent hypoxemia, hypercarbia, and apnea, often necessitating stepwise escalation of respiratory support [3].

Conventional bedside assessment during this critical period relies on a combination of maturity indices (e.g., gestational age), rapidly changing blood gas parameters, and clinical signs. Although informative, these parameters are inherently reactive, are influenced by ongoing interventions, and fail to provide an integrated, real-time assessment of the three core physiological pillars of respiratory competence: lung aeration, respiratory muscle performance, and pulmonary circulatory adaptation [4]. Consequently, clinicians lack a tool capable of integrating these dimensions into a comprehensive, early risk profile, potentially leading to delayed escalation of respiratory support in some infants and unnecessary intervention in others.

Point-of-care ultrasound (POCUS) offers a potential solution to this diagnostic gap by providing direct, non-



invasive visualization of cardiorespiratory physiology [5]. Individually, its components have proven clinical utility. The lung ultrasound score (LUS) reliably quantifies loss of lung aeration, predicts surfactant need, and identifies phenotypes of respiratory distress syndrome [6,7,8]. Diaphragmatic ultrasound can objectively assess respiratory muscle function by measuring diaphragmatic excursion and thickening fraction, parameters linked to extubation success and failure of noninvasive ventilation [9,10,11]. Furthermore, Doppler assessment of pulmonary blood flow, specifically the pulmonary acceleration time-to-ejection time (AT/ET) ratio, serves as a surrogate marker of pulmonary vascular resistance and circulatory adaptation during the transitional period [12,13].

Despite these advances, significant gaps continue to limit the clinical translation of ultrasound into a robust early predictive tool. First, most studies have primarily focused on single parameters, most commonly LUS, to predict outcomes such as surfactant requirement or continuous positive airway pressure failure [14,15]. Such studies typically report only discriminatory ability (e.g., area under the curve [AUC]) while overlooking essential metrics for clinical implementation, such as calibration (the agreement between predicted and observed risks) and clinical net benefit via decision curve analysis. Without such evaluations, it remains unclear whether these models can reliably inform threshold-based clinical decisions aimed at optimizing intervention timing. Second, although studies on diaphragm function [16] and transitional hemodynamics [17] have been reported, the available evidence remains fragmented across different patient cohorts and postnatal time points [18]. There is a critical lack of prospective data simultaneously quantifying lung aeration, diaphragmatic mechanics, and pulmonary circulation within a standardized, early postnatal window, precisely when most support escalations occur and before therapeutic interventions confound the baseline physiological state [19].

Therefore, there is an urgent need in the NICU for a reproducible early bedside tool that integrates these multiparametric physiological signals into an interpretable risk estimate, offering tangible clinical decision-making value to guide monitoring intensity and the timing of respiratory support escalation.

To address this need, we conducted a prospective observational cohort study in preterm infants (26+0 to 33+6 weeks' gestation). We hypothesized that a multiparametric ultrasound protocol performed within 6 hours after birth and integrating LUS, diaphragmatic excursion and thickening fraction, and pulmonary artery AT/ET ratio, would provide superior prediction of respiratory support escalation within the first 72 hours of life compared with models based on clinical variables alone or combined with LUS only. Our primary aim was to develop and compare these three incremental prediction models and to rigorously evaluate their discriminative performance, calibration, and clin-

ical net benefit. Secondary aims were to verify the inter-observer reproducibility of the ultrasound measurements and to establish a feasible early risk-stratification approach that translates the structural, mechanical, and circulatory aspects of pulmonary immaturity into actionable bedside information.

2. Materials and Methods

2.1 Study Design and Setting

This was a single-center prospective observational cohort study conducted in a tertiary NICU. The study period was from January 1, 2024, to June 30, 2025. The protocol adhered to a noninterventional design; all ultrasound examinations were performed solely for research purposes and did not influence clinical management. The study was approved by the Institutional Review Board of Shijiazhuang No.4 Hospital, Hebei Province, China (Approval No. 20220085). Written informed consent was obtained from the legal guardian of each infant prior to participation.

Clinical management followed a standardized departmental protocol for preterm infants, which included predefined oxygen saturation targets (90–95%), standardized noninvasive ventilation settings, and explicit criteria for surfactant administration and intubation. This approach minimized the potential for variations in clinical practice to systematically bias the observation of the study outcome.

2.2 Study Population

2.2.1 Inclusion and Exclusion Criteria

Eligible participants were preterm infants with a gestational age between 26+0 and 33+6 weeks who were admitted to the NICU immediately after birth. A key eligibility criterion was the ability to complete the full multiparametric ultrasound assessment within ≤ 6 hours after birth.

Infants were excluded if they met any of the following criteria: primary congenital heart disease (CHD) (excluding patent ductus arteriosus [PDA] or patent foramen ovale [PFO]); severe congenital malformations of the chest wall or diaphragm; pneumothorax requiring emergent surgical intervention within the first 6 hours; redirection of care to comfort measures only within 6 hours after birth; withdrawal of consent; or severely limited acoustic windows precluding completion of key ultrasound measurements.

2.2.2 Participant Enrollment and Flow

Eligibility was assessed by the clinical team once the infant was hemodynamically stable. Following guardian-informed consent, the baseline ultrasound examinations and collection of clinical data were completed. Participants were followed for 72 hours for assessment of the primary outcome and until hospital discharge for exploratory outcomes.

The primary analysis set included all infants who completed the baseline ultrasound within ≤ 6 hours after birth and for whom the primary outcome was ascertained at 72

hours. A prespecified reproducibility subset comprising the first 30 consecutively enrolled infants was used for inter- and intraobserver consistency analysis.

2.3 Sample Size Estimation

The primary outcome was the escalation of respiratory support within 72 hours after birth. Based on local historical data, the estimated incidence of this outcome was 40%. For the most complex predictive model (Model C), we anticipated a maximum of 8 candidate predictors. Following the recommendation of at least 10 outcome events per predictor variable, a minimum of 80 outcome events was required, corresponding to a total sample size of 200 infants. A sample size calculation based on comparisons of AUC further supported this estimate. Assuming an AUC of 0.75 for the intermediate model (Model B) and 0.83 for the full model (Model C), with a correlation of 0.60 between the receiver operating characteristic (ROC) curves, a sample size of 200 infants provided 80% power at a two-sided α of 0.05. Anticipating a 5% rate of attrition or missing data, we planned to enroll 210 infants over the 18-month study period.

2.4 Outcome and Predictor Definitions

2.4.1 Primary Outcome

The primary outcome was the escalation of respiratory support within the first 72 hours after birth. Respiratory support was categorized using an ordinal scale: Grade 0 (room air); Grade 1 (low-flow nasal cannula, <2 L/min); Grade 2 (high-flow nasal cannula, ≥ 2 L/min); Grade 3 (continuous positive airway pressure); Grade 4 (noninvasive positive pressure ventilation); Grade 5 (conventional invasive mechanical ventilation); and Grade 6 (high-frequency oscillatory ventilation). Escalation was defined as the first occurrence of either: (1) an increase by at least one grade on this scale, or (2) the administration of surfactant while the infant remained nonintubated (i.e., on noninvasive support). This definition established a composite endpoint capturing both increased support intensity and a significant therapeutic intervention.

2.4.2 Clinical Variables and Escalation Triggers

Prespecified clinical predictor variables for model construction included: gestational age (determined by first-trimester ultrasonography), male sex, completion of a full course of antenatal corticosteroids (defined as two doses of betamethasone administered 24 hours apart), and the 5-minute Apgar score. Birth weight was recorded as a baseline characteristic but was not included in the prediction models due to its high collinearity with gestational age and to preserve statistical power.

The clinical team made decisions regarding respiratory support escalation according to a standardized unit protocol. Objective trigger criteria included: (1) a requirement for a fractional inspired oxygen (FiO_2) ≥ 0.35 to maintain target saturations (90–95%) for at least 30 minutes; (2) arte-

rial partial pressure of carbon dioxide (PaCO_2) ≥ 65 mmHg with a pH ≤ 7.20 ; or (3) two or more apnea episodes within 1 hour requiring positive-pressure ventilation. The research team recorded these events and their timing without influencing clinical management.

2.5 Multiparametric Ultrasound Protocol

All ultrasounds were performed at the bedside within ≤ 6 hours of birth using a portable ultrasound system. Infants were examined in the supine position in a thermoneutral environment. Prewarmed gel was used, and gentle swaddling was employed as needed to minimize agitation; no sedation was administered. Ultrasound examinations were conducted during the infant's prevailing clinical condition. To reflect real-world bedside practice, no adjustments were made to existing respiratory support during the examination.

2.5.1 Lung Ultrasound Score (LUS)

Lung aeration was assessed using a standardized six-zone protocol, consisting of the anterior upper, anterior lower, and lateral lung zones on each side. Each zone was scored as follows: 0, regular aeration: A-lines with lung sliding; 1, moderate loss: multiple, separated B-lines covering $<50\%$ of the zone; 2, severe loss: confluent B-lines or "white lung" covering $\geq 50\%$; and 3, consolidation. Scores from all six zones were calculated to obtain a total LUS, ranging from 0 to 18. The most severe ultrasound pattern observed within each zone determined its score [6].

2.5.2 Diaphragm Ultrasound

Diaphragm function was assessed on the right side, with the left side examined when the right hemidiaphragm was inaccessible. Diaphragmatic excursion was measured in the coronal plane below the costal margin using M-mode ultrasonography, and the peak-to-peak displacement was averaged across three quiet respiratory cycles. Diaphragmatic thickening fraction was measured at the zone of apposition in the midaxillary line (8th–9th intercostal space) using a linear probe. Diaphragm thickness was measured at end-expiration and end-inspiration, and the thickening fraction was calculated as follows: $[(\text{inspiratory thickness} - \text{expiratory thickness}) / \text{expiratory thickness}] \times 100\%$. The mean value of three measurements was used for analysis.

2.5.3 Pulmonary Artery Doppler

Pulmonary blood flow dynamics were assessed from the parasternal short-axis view at the level of the aortic valve. Using a phased-array transducer, a pulsed-wave Doppler sample volume was placed 1–2 mm below the pulmonary valve, with an insonation angle $\leq 20^\circ$. The AT and ET were measured from stable Doppler waveforms over five cardiac cycles, and the AT/ET ratio was then calculated. Cardiac cycles with arrhythmia or imaging artefact were excluded.

2.5.4 Operator Training and Blinding

All examinations were performed by two sonologists after completing standardized training, which required demonstration of competency in ≥ 50 neonatal lung and diaphragm ultrasounds and ≥ 30 pulmonary artery Doppler studies. To minimize bias, a strict blinding protocol was maintained throughout the study. First, separation of roles was maintained: the two trained sonologists who performed the bedside ultrasound examinations were not involved in any clinical decision-making regarding respiratory support. Second, ultrasound findings were withheld from the clinical team: all ultrasound images were stored directly in a secure, deidentified research archive without any quantitative measurements or interpretations entered into the electronic medical record or communicated to the treating clinical team. No ultrasound reports were generated for clinical use. Third, delayed independent quantification was performed: all offline quantitative measurements (LUS, diaphragm excursion, diaphragm thickening fraction, AT/ET ratio) were performed by two independent readers blinded to all clinical data and infant outcome. These readers accessed only the deidentified ultrasound images from the research archive. All measurements were recorded in a dedicated research database separate from the hospital information system. This separation between image acquisition, clinical care, and quantitative analysis ensured that ultrasound findings did not influence escalation decisions during the 72-hour outcome period.

2.6 Data Collection and Quality Assurance

Data were collected prospectively using an electronic case report form with built-in range checks and dual-entry verification. All ultrasound cine loops and still images were stored in a deidentified archive. For the reproducibility subset, measurements were independently performed by two readers, and one reader repeated the measurements after a minimum interval of two weeks.

A dedicated quality control officer performed monthly audits on 10% of randomly selected cases. Prespecified criteria for excluding poor-quality measurements included incomplete cardiac cycles, excessive Doppler angle ($>20^\circ$), and significant artifact obscuring anatomical boundaries.

2.7 Statistical Analysis

2.7.1 Descriptive and Comparative Statistics

Continuous variables are presented as mean \pm standard deviation (SD) or median (interquartile range, IQR) based on data distribution, whereas categorical variables are presented as counts (percentages). Baseline characteristics were compared between the escalation and non-escalation groups using Student's *t*-test, Mann-Whitney U test, chi-square test, or Fisher's exact test, as appropriate. A two-sided *p*-value < 0.05 was considered statistically significant.

2.7.2 Prediction Model Development and Evaluation

Three sequential binary logistic regression models were developed to predict the primary outcome. Model A included baseline clinical variables: gestational age, neonatal sex, completion of antenatal corticosteroid therapy, and the 5-minute Apgar score. Model B added the LUS to the clinical variables. Model C, the complete multiparametric model, further included diaphragmatic excursion, diaphragmatic thickening fraction, and the pulmonary artery AT/ET ratio. All continuous variables were entered in their original scale, and no post hoc variable selection was performed. Potential interaction terms between predictors were not explored in these primary, prespecified models to preserve simplicity and clinical interpretability.

Model performance was assessed across three key domains. First, discrimination was quantified using the AUC, with 95% confidence intervals (CIs) derived from 1000 bootstrap samples. The DeLong test was used for pairwise AUC comparisons. Second, calibration, defined as the agreement between predicted probabilities and observed outcomes, was evaluated using calibration plots, calibration slope, and the Brier score. Third, the clinical utility of each model across a range of decision thresholds (10–50%) was evaluated using decision curve analysis, comparing the net benefit of each model against strategies of escalating respiratory support in all infants or in none.

To account for potential overfitting, Model C underwent internal validation using 1000 bootstrap resamples to calculate optimism-corrected performance estimates. Multicollinearity was assessed and deemed acceptable, with all variance inflation factors (VIF) below 2.5.

2.7.3 Internal Validation, Reproducibility, and Missing Data

Internal validation for the full model (Model C) was performed using 1000 bootstrap resamples to calculate optimism-corrected performance metrics, including the AUC and calibration slope. Multicollinearity among predictors was assessed using VIFs; all VIFs were < 2.5 , indicating no significant multicollinearity (see **Supplementary Table 1**).

Interobserver and intraobserver reproducibility for the key ultrasound measurements (LUS, diaphragmatic excursion, diaphragmatic thickening fraction, and AT/ET ratio) was assessed in the prespecified subset of 30 infants using a two-way random-effects, absolute-agreement intraclass correlation coefficient (ICC).

Missing data were handled according to a prespecified analysis plan. When the proportion of missing data for any key predictor was $\leq 15\%$, multiple imputation by chained equations (MICE) was applied using the mice package (version 4.3.2; R Foundation for Statistical Computing, Vienna, Austria) in R. The imputation model included all variables listed in Tables 1,2 (demographic, clinical, and ultrasound predictors) to preserve the underlying relationships within

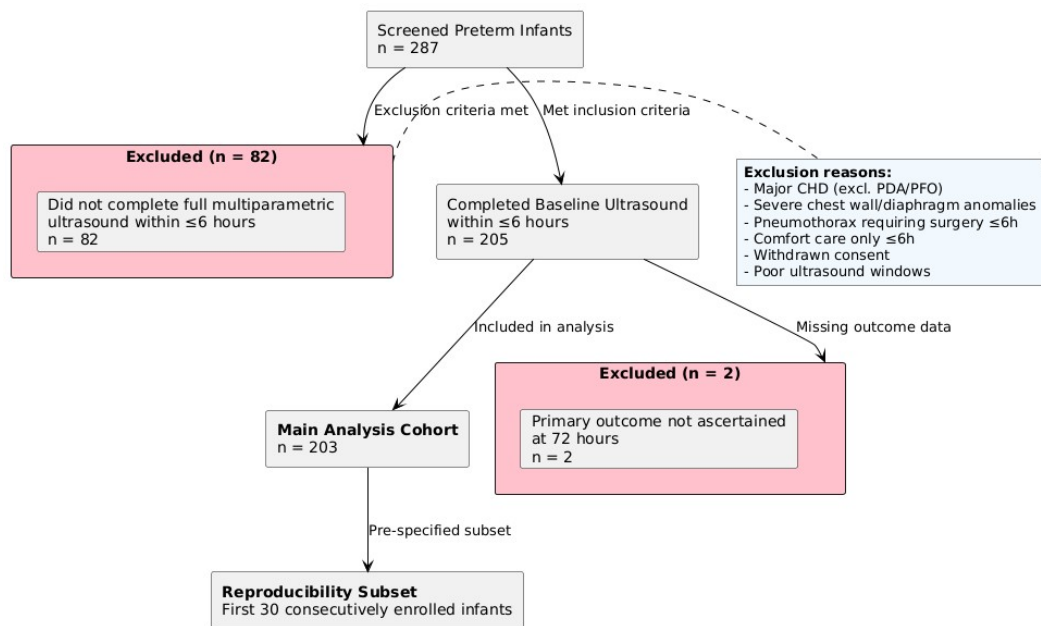


Fig. 1. Flow diagram of participant screening, enrollment, and inclusion in the different analyses phases. CHD, congenital heart disease; PDA, patent ductus arteriosus; PFO, patent foramen ovale.

the dataset. Predictive mean matching (PMM) was used for continuous variables (gestational age, 5-minute Apgar score, total LUS, diaphragmatic excursion, diaphragmatic thickening fraction, and AT/ET ratio). Logistic regression imputation was applied to binary variables (male sex, completion of a full course of antenatal corticosteroids). Ten imputed datasets were created, and model estimates were pooled using Rubin's rules. An overview of missing data is provided in **Supplementary Table 2**. All statistical analyses were performed using R software (version 4.3.2).

3. Results

3.1 Participant Flow and Characteristics

During the study period, 287 preterm infants were screened for eligibility. Of these, 205 infants completed the full multiparametric ultrasound assessment within ≤ 6 hours after birth. The primary outcome (respiratory support escalation within 72 hours) could not be ascertained for 2 infants, resulting in a primary analysis cohort of 203 infants (70.7% of screened infants). The first 30 consecutively enrolled infants formed the prespecified reproducibility subset (Fig. 1).

The baseline characteristics of the cohort stratified by the primary outcome are presented in Table 1. Infants who required escalation of respiratory support ($n = 82$, 40.4%) had significantly lower median gestational age (29.8 vs. 31.7 weeks, $p < 0.001$) and birth weight (1294 vs. 1618 g, $p = 0.002$). They were also less likely to have received a complete course of antenatal corticosteroids (63.4% vs. 83.5%, $p = 0.002$). In addition, the escalation group had higher initial PaCO₂ levels (58 vs. 47 mmHg, $p < 0.001$),

a higher prevalence of hemodynamically significant PDA (hsPDA) (28.0% vs. 13.2%, $p = 0.011$), and more frequent administration of postnatal caffeine (72.0% vs. 56.2%, $p = 0.027$). No significant differences were observed between groups regarding sex, 5-minute Apgar score, mode of delivery, or multiple gestation status.

Data on surfactant administration prior to the ≤ 6 -hour ultrasound assessment were not systematically recorded and therefore could not be reported.

Baseline values for the multiparametric ultrasound measurements, obtained within 6 hours of birth, are summarized in Table 2. The median LUS was 11 (IQR, 8–13), the median diaphragmatic excursion was 4.7 (IQR, 3.9–5.6) mm, the median diaphragmatic thickening fraction was 27.6% (IQR, 21.3–33.8), and the median pulmonary artery AT/ET ratio was 0.31 (IQR, 0.27–0.36).

3.2 Reproducibility of Ultrasound Measurements and Data Completeness

The reproducibility of the key ultrasound measurements was excellent. In the subset of 30 infants, all interobserver and intraobserver ICCs were ≥ 0.87 , indicating good-to-excellent agreement (Table 3).

Overall data completeness was high. Minor missing data ($< 5\%$) occurred for the 5-minute Apgar score ($n = 4$) and antenatal corticosteroid status ($n = 2$), and these variables were handled via multiple imputation (**Supplementary Table 1**). PaCO₂ values were missing in 9 cases (4.4%); however, because PaCO₂ was not included as a predictor in the final models, these values were not imputed.

Table 1. Baseline clinical and demographic characteristics.

Variable	Overall (n = 203)	Non-Escalation Group (n = 121)	Escalation Group (n = 82)	p-value
Gestational age (weeks)	30.9 [29.6–32.2]	31.7 [30.6–32.9]	29.8 [28.5–31.0]	<0.001
Birth weight (g)	1489 [1246–1718]	1618 [1422–1836]	1294 [1083–1516]	0.002
Male sex	108 (53.2)	58 (47.9)	50 (61.0)	0.085
Full course of ACS	153 (75.4)	101 (83.5)	52 (63.4)	0.002
5-min Apgar score	8 [7–9]	8 [8–9]	8 [7–9]	0.061
Cesarean delivery	138 (68.0)	80 (66.1)	58 (70.7)	0.541
Multiple gestation	46 (22.7)	24 (19.8)	22 (26.8)	0.305
First PaCO ₂ (mmHg)	52 [46–59]	47 [41–53]	58 [53–64]	<0.001
hsPDA present	39 (19.2)	16 (13.2)	23 (28.0)	0.011
Postnatal caffeine	127 (62.6)	68 (56.2)	59 (72.0)	0.027
In-hospital mortality, n (%)	6 (3.0)	-- †	-- †	-- †

Data presented as n (%) or median [IQR]. † In-hospital mortality was recorded for the cohort. A stratified analysis between escalation groups was not a prespecified endpoint of this early-prediction study. IQR, interquartile range; hsPDA, hemodynamically significant PDA.

Table 2. Baseline multiparametric ultrasound measurements (≤6 hours post-birth).

Variable	Full cohort (n = 203)	Unit/Scale
LUS total score	11 (8–13)	Points (0–18)
Diaphragmatic excursion	4.7 (3.9–5.6)	mm
Diaphragmatic thickening fraction	27.6 (21.3–33.8)	%
Pulmonary artery AT/ET ratio	0.31 (0.27–0.36)	Ratio

Data presented as median (IQR). LUS, lung ultrasound score; AT/ET, acceleration time-to-ejection time.

Table 3. ICC for ultrasound measurements.

Measurement	Interobserver ICC (95% CI)	Intraobserver ICC (95% CI)
LUS total score	0.94 (0.89–0.97)	0.97 (0.93–0.99)
Diaphragmatic excursion	0.91 (0.84–0.95)	0.94 (0.89–0.97)
Diaphragmatic thickening fraction	0.87 (0.78–0.93)	0.92 (0.85–0.96)
Pulmonary artery AT/ET ratio	0.89 (0.81–0.94)	0.93 (0.87–0.96)

ICC ≥0.75 indicates good reliability. ICC, intraclass correlation coefficient; CI, confidence interval.

3.3 Performance of the Predictive Models

3.3.1 Model Predictors and Discrimination

The results of the multivariable binary logistic regression analysis for the full model (Model C) are shown in Table 4. Lower gestational age, incomplete antenatal corticosteroid treatment, higher LUS, lower diaphragmatic excursion, a lower diaphragmatic thickening fraction, and lower AT/ET ratio were all independently associated with a higher risk of respiratory support escalation.

To illustrate the clinical application of the full multiparametric model, three representative cases spanning the spectrum of predicted risk are presented in **Supplementary Table 3**.

The discriminative performance of the three models is shown in Fig. 2A. Model C (clinical + LUS + diaphragmatic ultrasound parameters + AT/ET ratio) demonstrated significantly superior discriminative ability compared with both Model A (clinical variables only) and Model B (clinical variables + LUS), achieving an AUC of 0.88 (95% CI:

0.83–0.93), compared with 0.74 and 0.82, respectively (DeLong test, both $p < 0.05$).

3.3.2 Model Calibration and Diagnostics

The calibration plot for Model C showed excellent agreement between predicted probabilities and observed outcomes in both the apparent analysis and after bootstrap correction (Fig. 2B). This finding was confirmed by the quantitative metrics: the calibration slope was 0.99 (in the apparent analysis) and 0.97 (bootstrap-corrected), while the Brier score was 0.13, indicating low overall prediction error (Table 5). No evidence of significant multicollinearity was observed, with all VIFs well below 5 (**Supplementary Table 2**).

3.4 Clinical Utility: Decision Curve Analysis

The clinical utility of the models was assessed using decision curve analysis (Fig. 3). Across a clinically relevant range of threshold probabilities (approximately 0.22

Table 4. Multivariable logistic regression analysis for the full prediction model (Model C).

Predictor	Adjusted Odds Ratio (aOR)	95% CI	p-value
Gestational age (per 1 week)	0.79	0.67–0.94	0.006
Male sex (yes vs. no)	1.29	0.81–2.08	0.285
Full course of ACS (yes vs. no)	0.63	0.41–0.97	0.036
5-min Apgar (per 1 point)	0.91	0.80–1.04	0.162
LUS total score (per 1 point)	1.22	1.11–1.34	<0.001
Diaphragmatic excursion (per 1 mm)	0.73	0.60–0.89	0.002
Diaphragmatic thickening fraction (per 10%)	0.82	0.70–0.96	0.013
Pulmonary artery AT/ET ratio (per 0.01)	0.94	0.91–0.98	0.004

ACS, antenatal corticosteroids.

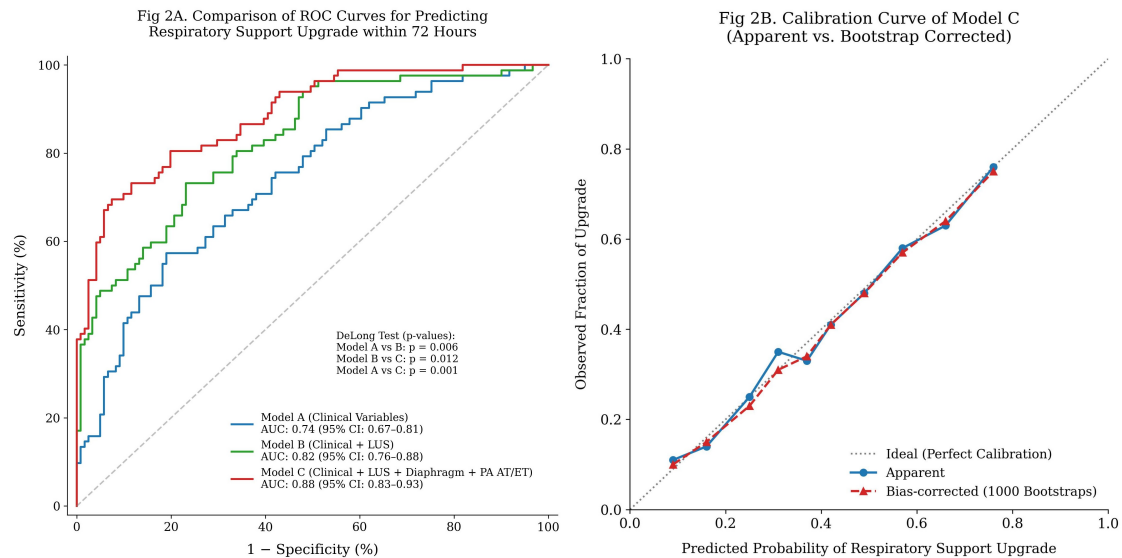


Fig. 2. Model performance evaluation. (A) ROC curves for the three prediction models. (B) Calibration plot for the full prediction model (Model C), showing apparent and bootstrap-corrected calibration. ROC, receiver operating characteristic.

Table 5. Performance metrics for prediction models.

Performance metric	Model A (Clinical)	Model B (Clinical + LUS)	Model C (Full Prediction Model)	Model C (Bootstrap-Corrected)
Brier score	0.19	0.16	0.13	0.14
Calibration intercept	-0.02	-0.01	-0.01	--
Calibration slope	0.93	0.97	0.99	0.97

The ideal calibration intercept is 0 and the ideal calibration slope is 1. Lower Brier score indicates better overall accuracy.

to 0.50), the full multiparametric model (Model C) provided a higher net benefit for clinical decision-making than both the simpler models (A and B) and the default strategies of “escalate all” or “escalate none”.

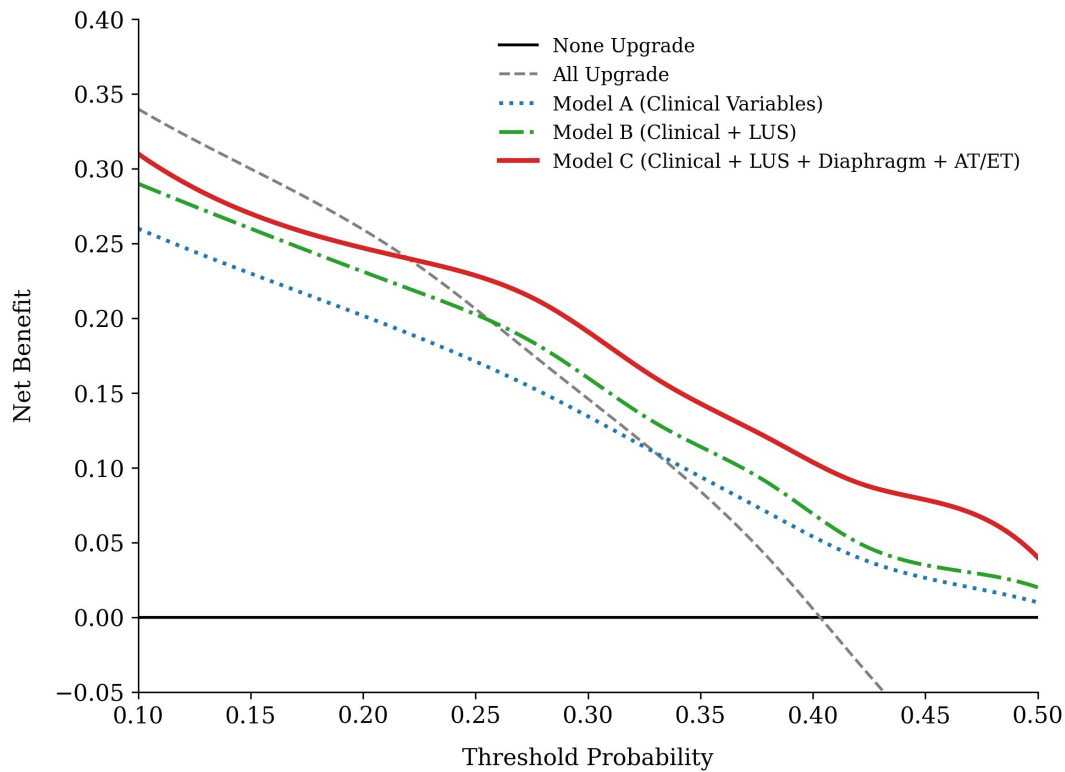
3.5 Exploratory In-Hospital Outcomes

Exploratory outcomes are summarized in Table 6. The median time to first respiratory support escalation was 7.8 hours (IQR, 3.4–18.6). The rates of noninvasive ventilation failure, repeated surfactant administration within 7 days, and in-hospital mortality were 15.3%, 8.9%, and 3.0%, respectively.

4. Discussion

4.1 Principal Findings

In this prospective cohort study, we demonstrated that a multiparametric ultrasound protocol performed within 6 hours after birth, integrating structural (LUS), mechanical (diaphragmatic ultrasound), and circulatory (pulmonary artery Doppler) assessments, significantly improves the early prediction of respiratory support escalation in preterm infants. The full model (Model C), which incorporated all three sonographic dimensions alongside key clinical variables, achieved superior discriminatory performance



This figure illustrates the relative net benefit for clinical decision-making and does not provide statistical significance testing.

Fig. 3. Decision curve analysis. Net benefit of each prediction model and default strategies across a range of threshold probabilities.

Table 6. Exploratory in-hospital outcomes.

Outcome	Events, n (%)	Time-related metric (Median [IQR])
Time to first escalation (hours)	82 (40.4)	7.8 [3.4–18.6]
Total duration of support (days)	203 (100)	6.2 [2.9–15.4]
Noninvasive ventilation failure	31 (15.3)	--
Repeated surfactant (within 7 days)	18 (8.9)	--
In-hospital mortality	6 (3.0)	--

(AUC: 0.88), excellent calibration, and provided a greater net clinical benefit across a range of decision thresholds compared with models based solely on clinical factors or on clinical variables combined with LUS alone. This performance, which remained robust after rigorous internal bootstrap validation, underscores that respiratory support escalation is a multifactorial event that is more accurately captured by a physiology-integrated approach than by maturity indices or lung aeration status alone [8].

4.2 Interpretation and Pathophysiological Context

Our findings are consistent with the evolving understanding of neonatal respiratory failure as a triad of impaired lung aeration, inefficient respiratory pump function, and altered pulmonary perfusion [4]. Although LUS effectively quantifies loss of lung aeration and interstitial fluid accumulation, which are key drivers of hypoxemia and surfactant deficiency [6,7], it does not fully assess

respiratory muscle reserve or pulmonary vascular adaptation. Diaphragm ultrasound provided independent prognostic value, as reduced diaphragmatic excursion and thickening fraction reflect impaired capacity to generate adequate tidal volume, thereby increasing the risk of hypercarbia and apnea, particularly in infants with highly compliant chest walls [10,16]. Concurrently, a lower pulmonary artery AT/ET ratio, indicative of elevated pulmonary vascular resistance, was independently associated with an increased risk of respiratory support escalation. This parameter likely reflects compromised pulmonary perfusion and ventilation-perfusion mismatch, thereby contributing to oxygenation failure that may not be directly detected by LUS alone [12,17]. Integration of these physiological signals creates a more comprehensive bedside assessment of pulmonary immaturity, which likely explains why Model C outperformed models lacking these components.

The independent associations of traditional risk factors, such as lower gestational age and incomplete antenatal corticosteroid treatment, with escalation risk confirm the fundamental role of structural immaturity and surfactant deficiency in neonatal respiratory failure [20]. Our model does not replace these critical clinical variables but rather complements them with real-time, modifiable physiological data, thereby potentially refining risk stratification among infants of similar gestational ages.

4.3 Comparison with Existing Literature and Clinical Implications

Current evidence supporting the use of ultrasound for respiratory prognostication remains fragmented. Many studies have validated LUS for predicting surfactant need or failure of noninvasive ventilation; however, these investigations have primarily reported discriminative metrics without adequately evaluating calibration or clinical utility [14,15]. Similarly, studies examining diaphragmatic function [9] or pulmonary hemodynamics [13] have typically been conducted independently or at different postnatal time points, thereby preventing an integrated assessment [18]. Our study addresses these gaps by prospectively quantifying all three physiological dimensions within a standardized early postnatal window (≤ 6 h after birth) and by employing a comprehensive model evaluation framework, assessing discrimination, calibration, and decision-curve net benefit, as recommended for clinical prediction models [21].

Our choice of the six-zone LUS protocol warrants further justification. Although extended protocols (e.g., 10- or 12-zone approaches) have been described and may provide more detailed granular regional aeration data, we selected the six-zone method for several reasons specific to our early postnatal assessment window. First, this protocol has been well-validated in preterm populations for predicting surfactant need and noninvasive ventilation failure [6,7], thereby providing a robust evidence base. Second, the six-zone approach offers a balance between comprehensiveness and efficiency; in our experience, it can be completed within 3–5 minutes, which is critical when assessing clinically unstable infants during the first 6 hours after birth. Third, minimizing handling and examination time reduces infant agitation and motion artifact, which likely directly contributed to the high measurement reproducibility observed in our study ($ICC \geq 0.94$). Finally, although more extensive protocols may offer incremental diagnostic information, our primary aim was to evaluate the additive value of a multiparametric assessment strategy (diaphragm + Doppler) beyond a practical and clinically applicable LUS baseline. Therefore, the six-zone protocol represents a pragmatic compromise between diagnostic detail and real-world feasibility in the time-sensitive NICU setting.

Our study employed trained sonologists to ensure measurement consistency during this validation phase. However, the protocol was designed for future clinical

translation. Standardized LUS systems are increasingly being performed by clinicians, and a recent systematic review concluded that LUS is feasible when performed by appropriately trained neonatologists [22]. Formal training curricula and credentialing pathways for neonatal POCUS, which encompasses lung, cardiac, and procedural applications, have now been established in North America, as evidenced by a recent national position statement [23]. The excellent interobserver reproducibility ($ICC \geq 0.87$) we observed further suggests that, with such structured competency-based training, reliable measurements are attainable by clinical NICU teams, supporting the potential for widespread application.

The high reproducibility of all ultrasound measurements ($ICC \geq 0.87$) confirms the feasibility of this protocol in unsedated neonates and supports its potential for broader clinical implementation [24]. Furthermore, the observed clustering of respiratory support escalation events early after birth (median, 7.8 hours) validates the clinical rationale for performing assessments within the ≤ 6 -hour assessment window, which aims to inform management decisions before significant therapeutic interventions.

4.4 Limitations and Future Directions

This study has several limitations. First, as a single-center study, our findings require external validation across diverse clinical settings with varying patient demographics, levels of ultrasound expertise, and respiratory management protocols (e.g., surfactant use strategies) [25]. Second, while we employed a strict blinding protocol and pre-defined objective escalation criteria, the observational design cannot fully exclude the possibility of residual confounding or subtle clinician-related bias in treatment decisions. Third, diaphragm ultrasound metrics can be influenced by the mode and level of concurrent respiratory support. Our study design prioritized a pragmatic bedside assessment approach; therefore, ultrasound examinations were performed without altering the infant's ongoing respiratory management. Consequently, our measurements reflect a combination of intrinsic muscle function and the effects of ongoing respiratory assistance rather than isolated muscle performance. However, this composite assessment likely represents the most clinically relevant context for a predictive tool intended for use in the dynamic NICU environment. Fourth, although the ≤ 6 -hour window was designed to capture early postnatal physiology, some infants may have received stabilizing support (e.g., continuous positive airway pressure [CPAP]) before ultrasound evaluation. Consequently, the ultrasound may partially reflect a postintervention physiological phenotype. Fifth, we used conventional logistic regression without exploring nonlinear relationships or interaction effects among predictors. Future work could investigate more flexible machine learning methods (e.g., random forest models) to enhance predictive performance. Sixth, we did not apply formal multiple

comparison corrections (e.g., Bonferroni) to the model performance metrics (e.g., AUC comparisons). Our analyses were based on prespecified sequential model comparisons rather than exploratory testing across unrelated hypotheses. Furthermore, internal validation via bootstrap resampling was performed to correct for overoptimism. The consistency of our findings across discrimination, calibration, and clinical utility metrics supports the robustness of the primary conclusion regarding the added value of the multiparametric model. Seventh, our primary outcome focused on short-term respiratory support escalation; therefore, the association between this model and longer-term outcomes, such as bronchopulmonary dysplasia, remains to be investigated.

Future research should prioritize external validation of the model and, if successful, progress to interventional trials to evaluate whether model-guided management strategies (e.g., targeting high-risk infants for enhanced monitoring or early selective therapy) improve clinical outcomes without increasing unnecessary interventions.

5. Conclusions

In conclusion, a bedside multiparametric ultrasound protocol performed within the first 6 hours of life, integrating lung aeration, diaphragmatic function, and pulmonary hemodynamics, significantly enhances the early prediction of respiratory support escalation in preterm infants compared with conventional clinical assessments or single-parameter imaging approaches. The model provides well-calibrated risk probabilities, demonstrates a positive net benefit for clinical decision-making, and is based on reproducible measurements. This approach translates the pathophysiology of pulmonary immaturity into a practical quantitative stratification tool and represents a promising strategy for optimizing early respiratory management in the NICU.

Availability of Data and Materials

All data generated or analyzed during this study are included in this manuscript.

Author Contributions

PW and LD: Conceptualization, PW, LD and LZ: Data curation, Formal analysis, YS and YW: Funding acquisition, Investigation, YS, YW, PW, and LZ: Methodology, Project administration, PW, LD and YW: writing – original draft, edit draft, YS, YW, and YC: Conception and design of the work, Review the draft. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

The study was carried out in accordance with the guidelines of the Declaration of Helsinki. The study was approved by the Institutional Review Board of Shijiazhuang No.4 Hospital, Hebei Province, China (Approval No. 20220085). Written informed consent was obtained from the legal guardian of each infant prior to participation.

Acknowledgment

The authors thank the nursing and medical staff of the Neonatal Intensive Care Unit at Shijiazhuang No.4 Hospital for their support in patient recruitment and data collection. We are grateful to the research assistants who contributed to data entry and quality assurance. We also acknowledge the infants and their families for their participation in this study.

Funding

This work was supported by the Hebei Medical Science Research Project (Grant Number 20231672).

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used DeepSeek (Version V3), Grammarly (Business Version 6.8.312) to check spelling and grammar. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/CEOG49489>.

References

- [1] Tana M, Tirone C, Aurilia C, Lio A, Paladini A, Fatore S, et al. Respiratory Management of the Preterm Infant: Supporting Evidence-Based Practice at the Bedside. *Children* (Basel, Switzerland). 2023; 10: 535. <https://doi.org/10.3390/children10030535>
- [2] Diggikar S, Ramaswamy VV, Koo J, Prasath A, Schmölder GM. Positive Pressure Ventilation in Preterm Infants in the Delivery Room: A Review of Current Practices, Challenges, and Emerging Technologies. *Neonatology*. 2024; 121: 288–297. <https://doi.org/10.1159/000537800>
- [3] Sweet DG, Carnielli VP, Greisen G, Hallman M, Klebermass-Schrehof K, Ozek E, et al. European Consensus Guidelines on the Management of Respiratory Distress Syndrome: 2022 Update. *Neonatology*. 2023; 120: 3–23. <https://doi.org/10.1159/000528914>
- [4] Sehgal A, Ruoss JL, Stanford AH, Lakshminrusimha S, McNamara PJ. Hemodynamic consequences of respiratory interventions in preterm infants. *Journal of Perinatology : Official*

Journal of the California Perinatal Association. 2022; 42: 1153–1160. <https://doi.org/10.1038/s41372-022-01422-5>

- [5] Recker F, Kipfmueller F, Wittek A, Strizek B, Winter L. Applications of Point-of-Care-Ultrasound in Neonatology: A Systematic Review of the Literature. *Life (Basel, Switzerland)*. 2024; 14: 658. <https://doi.org/10.3390/life14060658>
- [6] Brat R, Yousef N, Klifa R, Reynaud S, Shankar Aguilera S, De Luca D. Lung Ultrasonography Score to Evaluate Oxygenation and Surfactant Need in Neonates Treated With Continuous Positive Airway Pressure. *JAMA Pediatrics*. 2015; 169: e151797. <https://doi.org/10.1001/jamapediatrics.2015.1797>
- [7] Vardar G, Karadag N, Karatekin G. The Role of Lung Ultrasound as an Early Diagnostic Tool for Need of Surfactant Therapy in Preterm Infants with Respiratory Distress Syndrome. *American Journal of Perinatology*. 2021; 38: 1547–1556. <https://doi.org/10.1055/s-0040-1714207>
- [8] Zong H, Huang Z, Zhao J, Lin B, Fu Y, Lin Y, et al. The Value of Lung Ultrasound Score in Neonatology. *Frontiers in Pediatrics*. 2022; 10: 791664. <https://doi.org/10.3389/fped.2022.791664>
- [9] Bahgat E, El-Halaby H, Abdelrahman A, Nasef N, Abdel-Hady H. Sonographic evaluation of diaphragmatic thickness and excursion as a predictor for successful extubation in mechanically ventilated preterm infants. *European Journal of Pediatrics*. 2021; 180: 899–908. <https://doi.org/10.1007/s00431-020-03805-2>
- [10] Gupta R, Bandyopadhyay T, Yadav B, Guleria M, Maria A. Sonographic assessment of diaphragmatic thickness and excursion to predict CPAP failure in neonates below 34 weeks of gestation: A prospective cohort study. *Pediatric Pulmonology*. 2023; 58: 2889–2898. <https://doi.org/10.1002/ppul.26608>
- [11] Eskander AM, Abd-Elhameed AM, Osman NM, Magdy SM, ElKess GE. Role of diaphragmatic ultrasound in predicting weaning success from mechanical ventilation in pediatric intensive care unit. *Egyptian Journal of Radiology and Nuclear Medicine*. 2024; 55: 114. <https://doi.org/10.1186/s43055-024-01285-0>
- [12] Wang L, Zhang F, Li J, Liu Z, Kou Y, Song Y, et al. Using pulmonary artery acceleration time to evaluate pulmonary hemodynamic changes on preterm infants with respiratory distress syndrome. *Translational Pediatrics*. 2021; 10: 2287–2297. <https://doi.org/10.21037/tp-21-341>
- [13] de Boode WP, Singh Y, Molnar Z, Schubert U, Savoia M, Sehgal A, et al. Application of Neonatologist Performed Echocardiography in the assessment and management of persistent pulmonary hypertension of the newborn. *Pediatric Research*. 2018; 84: 68–77. <https://doi.org/10.1038/s41390-018-0082-0>
- [14] Alsina-Casanova M, Brito N, Balcells-Esponera C, Herranz-Barbero A, Teresa-Palacio M, Soler-García A, et al. Predictors of CPAP failure after less-invasive surfactant administration in preterm infants. *Frontiers in Pediatrics*. 2024; 12: 1444906. <https://doi.org/10.3389/fped.2024.1444906>
- [15] Dell’Orto V, Nobile S, Correani A, Marchionni P, Giretti I, Rondina C, et al. Early nasal continuous positive airway pressure failure prediction in preterm infants less than 32 weeks gestational age suffering from respiratory distress syndrome. *Pediatric Pulmonology*. 2021; 56: 3879–3886. <https://doi.org/10.1002/ppul.25678>
- [16] Dassios T, Vervenioti A, Dimitriou G. Respiratory muscle function in the newborn: a narrative review. *Pediatric Research*. 2022; 91: 795–803. <https://doi.org/10.1038/s41390-021-01529-z>
- [17] Chakkarapani AA, Roehr CC, Hooper SB, Te Pas AB, Gupta S, ESPR Neonatal Resuscitation section writing group. Transitional circulation and hemodynamic monitoring in newborn infants. *Pediatric Research*. 2024; 96: 595–603. <https://doi.org/10.1038/s41390-022-02427-8>
- [18] Martini S, Corsini I, Corvaglia L, Suryawanshi P, Chan B, Singh Y. A scoping review of echocardiographic and lung ultrasound biomarkers of bronchopulmonary dysplasia in preterm infants. *Frontiers in Pediatrics*. 2023; 11: 1067323. <https://doi.org/10.3389/fped.2023.1067323>
- [19] Abdallah Y, Mkony M, Noorani M, Moshiro R, Bakari M, Manji K. CPAP failure in the management of preterm neonates with respiratory distress syndrome where surfactant is scarce. A prospective observational study. *BMC Pediatrics*. 2023; 23: 211. <https://doi.org/10.1186/s12887-023-04038-6>
- [20] Htun ZT, Hairston JC, Gyamfi-Bannerman C, Marasch J, Duarte Ribeiro AP. Antenatal Corticosteroids: Extending the Practice for Late-Preterm and Scheduled Early-Term Deliveries?. *Children (Basel, Switzerland)*. 2021; 8: 272. <https://doi.org/10.3390/children8040272>
- [21] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Annals of Internal Medicine*. 2019; 170: W1–W33. <https://doi.org/10.7326/M18-1377>
- [22] Winkler M, Baik-Schmeditz N, Gruber V, Pichler G, Schwabegger B. Diagnostic applications of neonatologist-performed lung ultrasound in the delivery room in late preterm and term neonates: A systematic review of literature. *Paediatric Respiratory Reviews*. 2026; 58: 3–10. <https://doi.org/10.1016/j.prv.2026.01.001>
- [23] Narvey M, Stein N, Elsayed Y. Applications of point-of-care ultrasonography in the NICU: A Canadian perspective. 2025. Available at: <https://cps.ca/en/documents/position/POCUS> (Accessed: 30 June 2025).
- [24] Puskarcz-Gąsowska J, Kruczek P, Hożejowski R, Stefańska M, Błaż W, Sadowska-Krawczyńska I, et al. Lung Ultrasound Score in Neonatal RDS: Agreement between Sonography Expert and Neonatologists with Varying Levels of Experience. *Healthcare (Basel, Switzerland)*. 2024; 12: 1450. <https://doi.org/10.3390/healthcare12141450>
- [25] Ramaswamy VV, Bandyopadhyay T, Abiramalatha T, Pullattayil S AK, Szczapa T, Wright CJ, et al. Clinical decision thresholds for surfactant administration in preterm infants: a systematic review and network meta-analysis. *EClinicalMedicine*. 2023; 62: 102097. <https://doi.org/10.1016/j.eclinm.2023.102097>