

Online Registry for Patent Ductus Arteriosus in Very Preterm Infants in Shenzhen: Protocol for a Multicenter, Prospective, Observational Cohort Study

Jie Qi

Jinan University

Hui Wang

Jinan University

Hanni Lin

Shenzhen Luohu People's Hospital, Shenzhen, China

Xiaoli Zhao

Shenzhen Children's Hospital, Shenzhen, China

Rui Li

Shenzhen Longgang Maternal and Child Healthcare Hospital (Longgang Maternal and Child Research Institute of Shantou University Medical College), Shenzhen, China

Shuyu Luo

Jinan University

Guixuan Hong

Jinan University

Boshi Yu

Jinan University

Zihao Shen

Jinan University

Sien Huang

Jinan University

Suying Li

Jinan University

Liting Zhu

Jinan University

Guichao Zhong

Jinan University

Zhangbin Yu

yuzhangbin@126.com

Study protocol

Keywords: Patent ductus arteriosus, very preterm infant, haemodynamically significant PDA, online registry, multicenter cohort study, risk factors, neurodevelopmental outcome, predictive model, internal validation, biobank, protocol

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10970736/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Patent ductus arteriosus (PDA) is a common congenital cardiovascular condition in very preterm infants, with a reported incidence of up to 70%. Haemodynamically significant PDA (hsPDA) is associated with increased risks of intraventricular hemorrhage, bronchopulmonary dysplasia, necrotizing enterocolitis, and mortality. Systematic multicenter research on PDA in very preterm infants is scarce in China, and no dedicated online PDA registry exists either nationally or internationally. This study aims to establish a multicenter, prospective, observational cohort with a web-based registry to investigate the incidence, risk factors, management strategies, and outcomes of PDA in very preterm infants in Shenzhen, and to develop and internally validate a predictive model for hsPDA.

Methods

This multicenter, prospective, observational cohort study will be conducted across four tertiary hospitals in Shenzhen, China, from January 1, 2026 to December 31, 2028. Very preterm infants (gestational age < 32 weeks or birth weight < 1,500 g) with echocardiographically confirmed PDA will be enrolled. An online platform (Perinatal Cloud; <https://perinatalcloud.com/>) will serve as the central registry for standardized data collection. The primary objective is to develop and internally validate a machine learning-based predictive model for hsPDA. The target sample size of 300 infants with PDA was determined using the events-per-variable (EPV) criterion, assuming 10–15 candidate predictors and an hsPDA rate of approximately 50% among PDA cases, yielding ≥ 10 EPV. Internal validation will be performed via 10-fold cross-validation, with model performance assessed by area under the receiver operating characteristic curve (AUC), calibration plots, and Brier scores. Secondary objectives include determining PDA/hsPDA incidence, identifying risk factors, evaluating neurodevelopmental outcomes at 18–24 months corrected age, and exploring putative biomarkers. Biological specimens (blood, stool, urine, breast milk) will be collected and stored in a dedicated biobank. A nested case–control analysis will match PDA cases to non-PDA controls (1:2 ratio) on gestational age, birth weight, sex, and center.

Discussion

This study will establish the first dedicated online PDA registry in China, providing comprehensive, prospectively collected epidemiological data on PDA in very preterm infants. The development and internal validation of an hsPDA prediction model may enable early risk stratification and inform preventive strategies. The linked biobank will support future mechanistic and biomarker studies. Findings will be disseminated through peer-reviewed publications and conference presentations.

Background

With the advancement of perinatal medicine, the capacity to treat extremely and very preterm infants has steadily improved, and the gestational age at which preterm infants can be successfully resuscitated has progressively shifted downward, with concomitant increases in survival rates^[1, 2]. Nevertheless, preterm-associated morbidities and mortality continue to represent major challenges in perinatal medicine. Very preterm infants (gestational age < 32 weeks) account for a substantial proportion of admissions to the neonatal intensive care unit (NICU), and their morbidity and mortality rates are markedly higher than those observed in term infants^[3].

Patent ductus arteriosus (PDA) is a common congenital cardiovascular disorder among very preterm infants, with a reported incidence as high as 70%^[4, 5]. The diagnostic criteria for PDA are as follows: postnatal age > 72 hours, with evidence of a shunt between the aorta and the left pulmonary artery demonstrated by transthoracic color Doppler or spectral Doppler echocardiography. PDA is closely associated with major complications of prematurity, including respiratory distress syndrome, bronchopulmonary dysplasia (BPD), and necrotizing enterocolitis (NEC), and may exert a long-term influence on neurodevelopmental prognosis^[4, 6]. A cohort study indicates that a longer duration of significant PDA is associated with a significantly elevated risk of death or bronchopulmonary dysplasia (adjusted OR 1.37, 95% CI 1.03–1.82)^[7].

Haemodynamically significant PDA (hsPDA) is defined by the presence of one of the following criteria: (1) PDA haemodynamic alterations consistent with echocardiographic grade E3 or E4, together with fulfilment of three out of five designated parameters; or (2) PDA haemodynamic alterations meeting one parameter of grade E3 or E4, accompanied by any of the following clinical manifestations: baseline heart rate > 160 beats per minute, systolic murmur $\geq$ Grade 3/6 at the left sternal border, or mean arterial pressure below the normative value for the corresponding gestational age. The echocardiographic grading criteria for PDA haemodynamic alterations (E1–E4) are as follows: Grade E1 denotes the absence of transductal flow evidence on two-dimensional or color Doppler imaging. Grade E2 requires three of the following five parameters: ductus arteriosus internal diameter < 1.5 mm, peak transductal flow velocity (V_{max}) > 2 m/s, left atrial-to-aortic root ratio (LA:Ao ratio) < 1.5:1, isovolumetric relaxation time (IVRT) < 50 ms, and normal diastolic flow in the superior mesenteric artery. Grade E3 requires three of the following five parameters: ductus arteriosus internal diameter 1.5–3.0 mm, transductal flow V_{max} < 2 m/s, LA:Ao ratio 1.5–2.0:1, IVRT 50–60 ms, and reduced or absent diastolic flow in the superior mesenteric artery. Grade E4 requires three of the following five parameters: ductus arteriosus internal diameter > 3.0 mm, transductal flow V_{max} < 2 m/s, LA:Ao ratio > 2.0:1, IVRT > 60 ms, and reversal of diastolic flow in the superior mesenteric artery^[8].

The left-to-right shunt caused by hsPDA may result in systemic hypoperfusion and pulmonary overcirculation. Systemic hypoperfusion can compromise cerebral perfusion and contribute to reperfusion injury, thereby increasing the risk of intraventricular hemorrhage (IVH) and periventricular leukomalacia (PVL)^[9], with potential adverse consequences for neurodevelopmental outcomes^[10]. Pulmonary overcirculation may lead to feeding intolerance, heightened risk of NEC, pulmonary

hemorrhage, increased requirement for mechanical ventilation, and elevated risk of BPD [7, 11]. hsPDA may also induce lactic acidosis, augment pulmonary blood flow, cause a marked increase in left ventricular preload and cardiac dysfunction, and may present with manifestations such as left atrial or left ventricular enlargement [8]. Studies have identified gestational age, gestational hypertension, and chorioamnionitis as factors influencing the development of hsPDA in preterm infants [12, 13]. Although prophylactic indomethacin administration reduces the incidence of severe intracranial hemorrhage, it does not appear to significantly affect the composite outcome of mortality or major neonatal morbidities [5, 14].

International data reveal substantial variation in ductal patency rates across different gestational age strata. A systematic review of spontaneous ductal closure indicates that among infants born at < 28 weeks of gestation, approximately two-thirds retain a patent ductus at postnatal day 4 and only 41% achieve spontaneous closure by postnatal day 7, whereas up to 78% of infants born at more mature gestational ages achieve closure by day 7; closure rates are substantially lower among those with larger ductal diameters [15].

Among very preterm infants (gestational age < 32 weeks) and very low birth weight infants (birth weight < 1,500 g), echocardiographically confirmed PDA is identified in approximately two-thirds of infants [12], and the incidence is inversely related to both gestational age and birth weight [15, 16]. Institutional data from Shenzhen People's Hospital between 2020 and 2022 indicate a PDA incidence of approximately 21% among very preterm infants (defined as GA < 32 weeks or BW < 1,500 g); the lower rate relative to international estimates likely reflects the higher mean gestational age of the local cohort and a more restrictive echocardiographic screening protocol. The incidence of hsPDA among PDA cases in this cohort was approximately 45–55%.

The occurrence of PDA in preterm infants is associated with multiple factors, including gestational age, birth weight, neonatal respiratory distress syndrome (NRDS), sepsis, and chorioamnionitis [13]. Parameters such as invasive respiratory support, LA:Ao ratio, mode of delivery, duration of mechanical ventilation, and Apgar scores offer predictive value for the development of hsPDA [17, 18]. However, owing to a paucity of high-quality evidence, the optimal timing for intervention in PDA among very preterm infants remains a subject of ongoing debate within the neonatology community, and neither domestic nor international consensus guidelines or expert recommendations have been issued to date [19]. Therefore, identifying the risk factors for PDA and developing robust predictive models are of critical importance for reducing its incidence and associated morbidity.

Despite widespread recognition of the adverse impact of PDA and its complications on very preterm infants, few countries have established dedicated PDA-specific registry systems for the ongoing surveillance of its etiology and prognosis [20, 21]. To date, no online PDA registry has been implemented at the national level, either domestically or internationally. With advances in the treatment of preterm infants and the progressive shift toward resuscitation at earlier gestational ages, preterm birth cohort

studies are being established in various countries to investigate the incidence and risk factors of related complications^[20]. However, most existing databases were not specifically designed for PDA research, necessitating retrospective data analyses that may compromise the reliability and validity of the findings.

Systematic research on PDA in very preterm infants remains limited, and its long-term prognosis is not yet fully understood. In China, there is currently no formal online PDA registry, nor are there comprehensive studies addressing the etiology and prognosis of PDA. Existing investigations are predominantly single-center or retrospective in design, with insufficient inclusion of PDA-relevant variables and limited sample sizes^[22, 23].

The present study is designed as a multicenter, prospective, observational cohort investigation based on the Utstein model. Its primary aim is to systematically examine the risk factors, management strategies, treatment modalities, and clinical outcomes of PDA in very preterm infants within the Shenzhen metropolitan area, thereby elucidating the etiology, therapeutic effectiveness, and long-term prognosis of PDA, and providing an evidence-based foundation for clinical practice. Early identification of high-risk populations, together with regular screening and accurate diagnosis, is essential for reducing PDA-related morbidity and mortality and for improving overall outcomes^[24]. This study will analyze clinical data from infants with PDA, identify relevant risk factors, determine the incidence of PDA, and develop and internally validate predictive models to facilitate early risk recognition and preventive strategies. In parallel, the study will establish an online registry system and a biological specimen repository, laying a robust foundation for future in-depth investigations into PDA.

Study Objectives

Primary Objectives

1. To ascertain the incidence rates of PDA and hsPDA within the very preterm infant cohort in Shenzhen.
2. To identify risk factors for hsPDA and to develop and internally validate a predictive model for early identification of infants at high risk of developing hsPDA.
3. To evaluate neurodevelopmental outcomes in children diagnosed with PDA at a corrected age of 18 to 24 months.

Secondary Objectives

1. To determine the frequency of PDA-associated complications among very preterm neonates.
2. To explore potential etiological mechanisms and identify putative biomarkers linked to PDA pathogenesis using the linked biobank.

3. To compare the incidence of complications between infants with non-hsPDA and those with hsPDA.
4. To contrast neurodevelopmental status at corrected ages of 18 to 24 months between children with PDA and those with hsPDA.
5. To develop an exploratory long-term prognostic model for adverse neurodevelopmental outcomes at 18–24 months corrected age, incorporating early-life risk factors and PDA characteristics.

Methods/Design

Study Design

The proposed study is a multicenter, prospective, observational cohort investigation. Data collection is planned to run from January 1, 2026 through December 31, 2028. According to 2024 census figures, the annual number of live births in Shenzhen approximates 149,100. The study will recruit more than 300 very preterm neonates with a confirmed diagnosis of PDA across the neonatal intensive care units of four tertiary medical centers in Shenzhen. Each infant enrolled will be monitored for a 24-month period following hospital discharge.

Shenzhen People's Hospital functions as the Coordinating Center, bearing responsibility for the study's conceptualization, initiation, coordination, and administrative oversight. To augment the generalizability of the findings, three additional centers have been invited to participate. All participating institutions are tertiary medical facilities possessing substantial clinical expertise in PDA management, managing an approximate annual caseload of 100 very preterm infants with PDA, and maintaining established outpatient follow-up infrastructure.

At each site, three designated neonatologists will fulfill distinct roles as data abstractor, data validator, and specimen collection technician. The research consortium will convene periodic workshops to review progress, uphold quality benchmarks, and facilitate the integration of new study sites. Prospective centers will undergo a rigorous qualification assessment and receive comprehensive initial training and ongoing support.

The protocol is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement for cohort studies^[25]. The development and validation of predictive models will adhere to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines^[26, 27].

Patient and Public Involvement

The initial conceptualization and design of this research did not incorporate direct engagement with patient representatives or members of the general public. During the execution phase, the investigative

group plans to constitute a Community Advisory Panel, which will include parents of newborns as well as expectant parents, to gather feedback and informed perspectives. Findings generated from this study will be communicated to participating families and the wider community through peer-reviewed academic outlets, public health outreach activities, and scientific conference presentations.

Informed Consent

The project involves the acquisition of clinical data and biological materials from both maternal participants and their neonates. Study investigators will furnish parents or legally authorized surrogates with a thorough explanation of all relevant aspects of the research, and documented informed consent will be secured prior to the initiation of any study-specific procedures. All activities described herein are conducted in full compliance with the ethical provisions set forth in the Declaration of Helsinki.

Study Participants

The collaborative network for this study comprises the following tertiary-level institutions in Shenzhen: Shenzhen People's Hospital (affiliated as The Second Clinical Medical College of Jinan University and The First Affiliated Hospital of Southern University of Science and Technology); Longgang District Maternity & Child Healthcare Hospital of Shenzhen, which also hosts the Longgang Maternity and Child Institute of Shantou University Medical College; Shenzhen Children's Hospital; and Shenzhen Luohu People's Hospital.

Infants born at or transferred to any of the participating sites who receive a confirmed diagnosis of PDA during the enrollment window—spanning January 1, 2026 through December 31, 2028—will undergo screening for potential study inclusion. The decision regarding whether to enroll in or decline participation in this observational investigation shall in no way alter or affect the routine clinical management provided to the infant.

Inclusion and Exclusion Criteria

Inclusion Criteria

- Preterm infants with a gestational age < 32 weeks OR birth weight < 1,500 g.
- Born at or transferred to one of the participating study hospitals.
- Echocardiographically confirmed diagnosis of PDA (postnatal age > 72 hours).
- Provision of signed informed consent by a parent or legally authorized guardian.

Exclusion Criteria

- Presence of major congenital anomalies.
- Diagnosis of complex congenital heart disease.
- Identified inborn errors of metabolism or genetic syndromes.
- Death within 24 hours of NICU admission.
- Withdrawal of life-sustaining treatment by the family or explicit refusal of study enrollment.

Quality Assurance for Echocardiographic Assessment

Given the central role of echocardiography in defining both PDA diagnosis and hsPDA severity grade (E1–E4), and the complexity of the multi-parameter grading scheme employed in this study, formal inter-rater reliability assessment will be conducted prior to and during the enrollment period. All participating sonographers and interpreting cardiologists will complete a standardized training module incorporating representative echocardiographic recordings across the E1–E4 spectrum. Inter-rater agreement (Cohen's κ or Fleiss' κ for multiple raters) will be calculated at study initiation and at 12-month intervals thereafter; a $\kappa \geq 0.75$ will be considered indicative of acceptable agreement. Centers with $\kappa < 0.75$ will undergo retraining. All echocardiographic images will be archived digitally to permit central adjudication of ambiguous cases by the Coordinating Center.

Outcome Definitions

In-Hospital Outcomes

- Diagnosis of PDA; diagnosis of hsPDA; PDA necessitating surgical or transcatheter intervention; vital status at discharge (survival/death); growth pattern at discharge.
- Anthropometric assessments: weight, length, head circumference, inpatient growth velocity, and incidence of extrauterine growth restriction.

Follow-Up Outcomes

- Rate of attrition (loss to follow-up); post-discharge survival status.
- Change in weight-for-age, length-for-age, and head circumference-for-age z-scores at 18–24 months corrected age.

- Healthcare resource utilization between discharge and 18–24 months corrected age.
- Echocardiographic findings during the follow-up interval.
- Comprehensive neurodevelopmental assessment conducted at 18–24 months corrected age.

Definition of Severe Neurodevelopmental Impairment (NDI)

Severe neurodevelopmental impairment (NDI) is operationally defined by the documentation of any one of the following conditions: bilateral loss of vision; auditory deficit requiring the use of hearing aids; inability to walk 10 steps without assistance; a clinical diagnosis of cerebral palsy; and/or a total score below 75 on the Griffiths Development Scales-Chinese (GDS-C) version. The grading of NDI is further categorized into three tiers: Severe NDI, indicated by a Mental Development Index (MDI) and/or Psychomotor Development Index (PDI) score of less than 70; Mild NDI, corresponding to MDI and/or PDI values falling between 70 and 84; and Normal Neurodevelopmental Status, defined by MDI and PDI scores both equal to or exceeding 85.

Data Collection and Management

The online registry system is architected upon a cloud-based platform featuring functionalities for data entry, secure storage, query, and statistical analysis. The system employs multi-layered encryption protocols to safeguard data integrity and confidentiality. Authorized personnel at participating sites will input patient information via a secure login according to a standardized template. Embedded data validation rules will automatically flag anomalous or inconsistent entries, thereby promoting data accuracy and completeness.

Enrollment of participants will be carried out on a prospective basis, with the recruitment phase anticipated to conclude by the end of 2028. A distinctive alphanumeric code will be allocated to each collaborating institution and to every infant entered into the study. Clinical information and associated biospecimen storage identifiers will be collated by the registry coordinating center utilizing the Perinatal Cloud online database platform (Perinatal-Neonatal Collaborative Platform, <https://perinatalcloud.com/>). All data entry will be performed by trained and certified data abstractors. Data auditors will conduct routine quality control assessments and provide feedback. The system architecture permits timely data correction and updates, enabling real-time data export for quality assurance and statistical interrogation.

An oversight body designated as the Data Monitoring Committee (DMC) will be established, consisting of the Principal Investigators representing both the Coordinating Center and each affiliated study site. The DMC holds responsibility for ensuring that all study activities are conducted in strict compliance with the authorized protocol. In the event that protocol revisions are indicated, the DMC will formally present proposed amendments to the appropriate ethics review boards and may only proceed following receipt of written approval. The committee will convene at regular intervals to evaluate study progress

and to verify continued adherence to established procedures. Beyond these supervisory functions, the DMC is also charged with the identification and onboarding of additional participating institutions, distribution of study-related materials, coordination of training workshops, validation and interpretation of accumulated data, and preparation of scientific reports for publication. In parallel, an ethics committee operating independently from the investigative team and funding sources will carry out annual appraisals of study conduct and progress.

Core Data Elements

Data collection encompasses three broad domains: (1) demographic and clinical information (maternal demographics, antenatal history, delivery characteristics, neonatal resuscitation, inpatient therapies across all organ systems, and growth parameters); (2) PDA-specific variables (clinical presentation, comprehensive echocardiographic metrics, management strategies, and ductal closure status); and (3) follow-up assessment data (echocardiographic re-evaluation, growth, feeding, and standardized neurodevelopmental testing). A detailed enumeration of all data elements is provided in Additional File 1. Key PDA-specific echocardiographic surveillance metrics include: PDA internal diameter, PDA diameter to left pulmonary artery diameter ratio, LA:Ao ratio, transductal flow velocity and direction, IVRT, left ventricular output, superior mesenteric artery diastolic flow patterns, and mitral valve E/A ratio.

Biological Sample Collection and Processing

The study will entail the collection of umbilical cord blood or peripheral venous plasma, stool samples, urine specimens, and maternal breast milk from participants diagnosed with PDA. Blood samples will be centrifuged at low temperature promptly post-collection, with plasma aliquots transferred to -80°C freezers for long-term cryopreservation. Other biological matrices will be temporarily maintained at 4°C prior to transfer to -80°C for long-term archiving. All specimen containers will be clearly labeled with the precise postnatal date and time of collection.

Detailed records pertaining to specimen volume, collection timepoints, and planned storage duration will be documented both on the physical specimen container and within the linked online database. Informed consent for the collection and banking of these biospecimens will be obtained concurrently with study enrollment. Given the inherent challenges and potential iatrogenic anemia associated with phlebotomy in very preterm infants, blood sampling volumes will be rigorously minimized and restricted to the minimal quantity required for validated diagnostic assays. Specimen procurement will be performed by experienced NICU nursing staff, with subsequent processing and cryopreservation managed by dedicated research personnel.

Nested Case–Control Design for Matched Comparisons

A nested case–control analysis will be conducted to compare outcomes between PDA cases and non-PDA controls. Given that the anticipated PDA incidence in the source population is approximately 50%, a 1:4 matching ratio is unlikely to be feasible without compromising match quality. Accordingly, a 1:2 matching ratio will be targeted. Control infants will be matched to PDA cases on gestational age (± 1 week), birth weight (± 100 g), sex, and participating center. Matching will be performed using nearest-neighbor propensity score matching without replacement. Priority for selection as a control will be given to concurrently hospitalized very preterm infants who have not received a diagnosis of PDA. Biological specimens will be collected from control infants following the same protocols applied to PDA cases. The matched design will be used in secondary analyses comparing complication rates and neurodevelopmental outcomes between groups.

Sample Size Calculation

The target sample size was determined using the events-per-variable (EPV) criterion for multivariable logistic regression, consistent with the TRIPOD guidelines for prediction model development ^[26, 27]. The following assumptions underpin the calculation:

- The primary outcome for model development is hsPDA (E3/E4 grade). Based on institutional data from Shenzhen People’s Hospital (2020–2022), the incidence of hsPDA among PDA cases is approximately 45–55%. A conservative estimate of 50% is adopted.
- The predictive model will incorporate 10–15 candidate predictor variables, selected based on clinical relevance and prior literature (including gestational age, birth weight, antenatal corticosteroid exposure, chorioamnionitis, LA:Ao ratio, ductal diameter, and early respiratory support parameters).
- A minimum of 10 EPV is required for stable coefficient estimation in logistic regression. With 15 candidate predictors, a minimum of 150 hsPDA events is needed.
- With an estimated hsPDA rate of 50% among enrolled PDA cases, 300 PDA cases will yield approximately 150 hsPDA events, satisfying the $EPV \geq 10$ criterion.
- Accounting for an anticipated 20% rate of incomplete data or loss to follow-up at the primary analysis time point, the enrollment target is set at 300 infants with confirmed PDA.
- Across the four participating centers, with an aggregate annual volume of approximately 100 very preterm infants with PDA, a 3-year recruitment period (January 2026–December 2028) is expected to achieve the target enrollment.

For the long-term prognostic model (predicting severe NDI at 18–24 months corrected age), assuming an NDI rate of approximately 20–25% in the PDA cohort, the 300 enrolled infants will yield 60–75 NDI events, permitting the inclusion of 6–7 candidate predictors at the $EPV \geq 10$ threshold. We acknowledge

that the sample size for this secondary analysis is more limited; its results will be regarded as exploratory and hypothesis-generating.

Statistical Analysis

Descriptive statistics will be employed to summarize the baseline demographic and clinical characteristics of the study cohort, along with the incidence and distribution of PDA. Continuous variables will be presented as mean $\pm$ standard deviation (SD) or median with interquartile range (IQR), depending on the normality of distribution assessed by the Shapiro–Wilk test and visual inspection of histograms and Q–Q plots. Categorical variables will be summarized as frequencies and percentages. Intergroup comparisons for continuous variables will be performed using the independent samples t-test or the Mann–Whitney U test, as appropriate. Categorical comparisons will utilize the Chi-square (χ^2) test or Fisher’s exact test.

Survival Analysis

The Kaplan–Meier product-limit method will be applied to estimate survival probabilities and evaluate the impact of PDA on survival duration. The log-rank test will be used to assess differences in survival distributions between defined subgroups.

Cox Proportional Hazards Regression

Multivariable Cox regression models will be constructed to evaluate the independent effect of multiple covariates on time-to-event outcomes and to identify independent prognostic determinants. Covariates considered for inclusion will encompass gestational age, birth weight, PDA classification and dimensions, and therapeutic modality. The proportional hazards assumption will be tested using Schoenfeld residuals.

Multivariable Logistic Regression

This approach will be utilized to examine the association between PDA status and the occurrence of short-term complications, with results expressed as adjusted odds ratios (aOR) and corresponding 95% confidence intervals (CI).

Predictive Model Development and Internal Validation

The primary predictive modeling objective is to develop a model for early identification of infants at high risk for hsPDA, using clinical and echocardiographic variables available within the first 72 hours of life. A

systematic approach to model development and validation will be employed, consistent with TRIPOD guidelines [26, 27]:

- **Variable selection:** Candidate predictors will be selected based on clinical knowledge and univariable screening ($P < 0.10$ for retention). Collinearity will be assessed using variance inflation factors (VIF), with $VIF > 10$ indicating problematic multicollinearity.
- **Model algorithms:** A suite of machine learning algorithms will be evaluated, including logistic regression with L1 (LASSO) and L2 (ridge) regularization, random forests, gradient boosting machines (XGBoost), and support vector machines. Logistic regression will serve as the reference model given its interpretability and established use in clinical prediction.
- **Internal validation:** Model performance will be assessed via 10-fold cross-validation repeated 10 times to obtain stable estimates of predictive performance. This approach balances computational feasibility with robust performance estimation. For each algorithm, the following metrics will be reported: area under the receiver operating characteristic curve (AUC) for discrimination, calibration-in-the-large and calibration slope (with accompanying calibration plots), and the Brier score as an overall measure of predictive accuracy.
- **Model comparison:** Nested models will be compared using the likelihood ratio test; non-nested models will be compared via DeLong's test for correlated AUCs. Decision curve analysis will be performed to evaluate net benefit across clinically relevant threshold probabilities.
- **Sensitivity analyses:** The robustness of findings will be examined by (a) repeating analyses stratified by center, (b) excluding infants who died before day 7, and (c) using complete-case data alongside imputed datasets.
- **A secondary prognostic model** will be developed for the prediction of severe NDI at 18–24 months corrected age, using Cox proportional hazards or logistic regression as appropriate to the outcome distribution. This model will be considered exploratory.

Missing Data

Missing data patterns will be characterized using descriptive summaries and visualizations. The primary approach to handling missing covariate data will be multiple imputation using chained equations (MICE), generating $m = 20$ imputed datasets under the assumption of missing at random (MAR). Imputation models will include all outcome variables and candidate predictors. Sensitivity to departures from the MAR assumption will be assessed using pattern-mixture models. Complete-case analysis will be reported as a sensitivity analysis for comparison. K-nearest neighbors (KNN) imputation will be employed as a supplementary method to evaluate the robustness of conclusions across imputation strategies.

All statistical computations and data visualizations will be executed using R software, version 4.3.3 (R Foundation for Statistical Computing, Vienna, Austria). Key packages will include *mice* for multiple imputation, *caret* and *tidymodels* for machine learning, *rms* for regression modeling and calibration, and *pROC* for ROC analysis. A two-tailed P-value < 0.05 will be considered indicative of nominal statistical significance. Given the exploratory nature of several secondary analyses, no formal correction for multiplicity will be applied; however, 95% confidence intervals will be reported for all effect estimates to convey the precision of findings.

Discussion

While select nations have instituted disease-specific online registries to facilitate ongoing surveillance of risk factors and outcomes associated with preterm PDA^[20, 21], a dedicated online registry for PDA is currently non-existent in China^[22], and multicenter prospective studies in this domain are scarce. Through the execution of this multicenter, prospective, observational investigation, this study will furnish a systematic appraisal of the epidemiological characteristics and prognostic trajectory of PDA among very preterm infants in the Shenzhen region. It is anticipated that this endeavor will foster greater standardization and harmonization of diagnostic and therapeutic approaches to PDA management. To this end, the research consortium has pioneered the establishment of the Shenzhen Very Preterm Infant PDA Online Registry. To catalyze progress in PDA research and therapeutics, the study team intends to share curated de-identified datasets with collaborating research networks upon reasonable request, subject to approval by the Data Monitoring Committee and compliance with institutional data-sharing agreements.

A prior European multicenter study documented marked regional variation in PDA treatment, ranging from 10% to 39%, and a higher risk of bronchopulmonary dysplasia or death among treated infants (adjusted risk ratio 1.33, 95% CI 1.18–1.51), underscoring the strong interrelationship between PDA and a spectrum of adverse neonatal complications^[28]. Through multivariable analytical approaches and machine learning techniques, the present study will further clarify the independent contributions of gestational age, birth weight, and PDA magnitude to prognostic outcomes, thereby offering more precise guidance for clinical intervention^[17, 29]. The development of an internally validated predictive model for hsPDA, grounded in prospectively collected multicenter data, represents a key methodological contribution. Should the model demonstrate acceptable discrimination and calibration, it may be deployed as a clinical decision-support tool to guide the intensity of echocardiographic surveillance and inform early management decisions.

With respect to long-term prognosis, this study is designed to assess whether PDA and hsPDA elevate the risk of adverse short- and long-term outcomes in the very preterm population^[30]. By leveraging a prospective online registry infrastructure to support extended longitudinal follow-up, we seek to characterize associations between PDA severity and long-term cognitive, motor, and cardiovascular outcomes. Mining of the amassed biospecimen data holds the potential to unveil novel biomarkers of

critical importance for both risk stratification and therapeutic targeting ^[31]. This stands to not only deepen our etiological comprehension of PDA but also to delineate promising avenues for future mechanistic inquiry.

Implementation Considerations and Challenges

Several operational challenges are anticipated. First, ensuring data homogeneity across four centers with potentially differing clinical practices and echocardiographic protocols will require rigorous standardization. The pre-enrollment inter-rater reliability training and periodic re-assessment described in the Methods are designed to address this; additionally, a central adjudication mechanism for ambiguous echocardiographic classifications will be implemented. Second, the multi-parameter hsPDA grading scheme (E1–E4), while comprehensive, may prove challenging to apply consistently in busy clinical settings. To mitigate this, we have developed a standardized echocardiographic reporting template integrated into the Perinatal Cloud platform, with automated grade calculation based on entered parameters. Third, clinician treatment preferences—including the choice and timing of pharmacologic intervention—introduce potential confounding by indication. We will collect detailed data on the rationale for treatment decisions and will employ propensity score methods in secondary analyses to reduce this source of bias. Fourth, loss to follow-up at 18–24 months is a well-recognized challenge in neonatal cohort studies ^[32]. Our mitigation strategy integrates study visits with routine high-risk developmental follow-up appointments, supplemented by telephone and WeChat-based reminders. We will characterize the pattern of missing follow-up data and, where appropriate, employ inverse probability weighting to account for differential attrition.

Study Limitations

Notwithstanding its multicenter, prospective design, this observational study is subject to inherent limitations. First, the geographic restriction of recruitment to healthcare institutions within the Shenzhen metropolitan area may curtail the external validity and generalizability of the findings to broader, more heterogeneous populations with different demographic profiles and clinical practices. External validation of the predictive model in independent cohorts will be necessary before broader clinical application ^[33]. Second, the follow-up duration of 24 months corrected age permits assessment of early neurodevelopmental outcomes but is insufficient to capture later-emerging cognitive, behavioral, and academic difficulties, or long-term cardiovascular sequelae. Extended follow-up of this cohort is planned in subsequent study phases, contingent upon securing additional funding. Third, this protocol does not mandate a detailed comparative effectiveness analysis of specific PDA treatment modalities (e.g., ibuprofen vs. acetaminophen vs. expectant management). While treatment data will be collected, the non-randomized design and anticipated confounding by indication preclude causal inference regarding treatment effects. Fourth, despite being powered for the primary hsPDA prediction model, the sample size may be insufficient for robust subgroup analyses (e.g., by gestational age strata) or for the

exploratory prognostic model of NDI. Fifth, the reliance on the Griffiths Development Scales-Chinese (GDS-C) for neurodevelopmental assessment, while culturally appropriate, may limit direct comparability with international cohorts assessed using the Bayley Scales of Infant Development.

Conclusions

This investigation will furnish essential empirical data regarding the epidemiology and prognostic landscape of PDA in very preterm infants, while simultaneously establishing infrastructure—the online registry and biobank—for sustained research productivity. By enhancing our understanding of the risk architecture and long-term sequelae of PDA, the overarching objectives of this research are to facilitate early risk stratification, inform individualized management, and ultimately reduce PDA-associated morbidity and neurodevelopmental impairment. Future scholarly pursuits should prioritize external validation of the predictive model in geographically diverse cohorts, rigorous comparative evaluation of early interventional strategies through randomized designs, and extended longitudinal follow-up to characterize the lifelong health implications of PDA sustained during extreme prematurity, including potential repercussions on cardiovascular health in adulthood.

Abbreviations

ABR: Auditory Brainstem Response

AKI: Acute Kidney Injury

aOR: Adjusted Odds Ratio

ARDS: Acute Respiratory Distress Syndrome

AUC: Area Under the Receiver Operating Characteristic Curve

BAEP: Brainstem Auditory Evoked Potentials

BMI: Body Mass Index

BPD: Bronchopulmonary Dysplasia

CI: Confidence Interval

DMC: Data Monitoring Committee

ECG: Electrocardiography

EEG: Electroencephalography

EPV: Events Per Variable

GDS-C: Griffiths Development Scales-Chinese

GMDS: Griffiths Mental Development Scales

HIE: Hypoxic-Ischemic Encephalopathy

hsPDA: Haemodynamically Significant Patent Ductus Arteriosus

IVH: Intraventricular Hemorrhage

IVRT: Isovolumetric Relaxation Time

IQR: Interquartile Range

KNN: K-Nearest Neighbors

LA:Ao: Left Atrial to Aortic Root Ratio

LASSO: Least Absolute Shrinkage and Selection Operator

MAR: Missing At Random

MDI: Mental Development Index

MICE: Multiple Imputation by Chained Equations

MRI: Magnetic Resonance Imaging

NBNA: Neonatal Behavioral Neurological Assessment

NDI: Neurodevelopmental Impairment

NEC: Necrotizing Enterocolitis

NICU: Neonatal Intensive Care Unit

NPO: Nil Per Os (Nothing by Mouth)

NRDS: Neonatal Respiratory Distress Syndrome

OR: Odds Ratio

PDA: Patent Ductus Arteriosus

PDI: Psychomotor Development Index

PPHN: Persistent Pulmonary Hypertension of the Newborn

PVL: Periventricular Leukomalacia

ROP: Retinopathy of Prematurity

SD: Standard Deviation

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

VIF: Variance Inflation Factor

Vmax: Peak Transductal Flow Velocity

Declarations

Ethics approval and consent to participate

This study protocol has received formal approval from the Medical Ethics Committee of Shenzhen People's Hospital (Approval Number: Version 1.0, Approval Date: October 15, 2024). Prior to the commencement of any study-related activities, each participating center is required to obtain site-specific approval from its respective Institutional Review Board. All parents or legal guardians will be thoroughly apprised of the study's objectives, procedures, potential risks, and benefits, and will provide written informed consent prior to enrollment. This study will be conducted in accordance with the ethical principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study will be available from the corresponding author on reasonable request, subject to approval by the Data Monitoring Committee. De-identified individual participant data underlying published results, together with the data dictionary and statistical analysis plan, will be made available to qualified researchers who submit a methodologically sound proposal. Data will be deposited in the Perinatal Cloud online database platform (<https://perinatalcloud.com/>). Data sharing will be governed by institutional data transfer agreements and will comply with the Human Genetic Resources Administration of China (HGRAC) regulations.

Competing interests

The authors declare that they have no competing interests.

Funding

Guangdong Provincial Medical Science and Technology Research Fund Project B2025538

Shenzhen Clinical Medical Research Center 20220819113341005

Shenzhen Clinical Medical Research Center for Children ' s Health and Diseases szcrc2024_013

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

Conceptualization: G.Z., Z.Y.

Methodology: G.Z., Z.Y., J.Q., H.W., H.L., X.Z., R.L.

Software: G.H., B.Y., Z.S.

Validation: J.Q., H.W., H.L., X.Z., R.L.

Formal analysis: J.Q., Z.Y.

Investigation: S.H., S.L., L.Z.

Data curation: S.H., S.L., L.Z., G.H.

Writing – original draft: J.Q., H.W.

Writing – review & editing: All authors.

Visualization: G.H., B.Y.

Supervision: G.Z., Z.Y.

Project administration: G.Z., Z.Y.

Funding acquisition: Z.Y.

All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

The authors wish to thank all participating hospitals, the clinical and research staff at each study site, and the families who will participate in this study.

Authors' information

Jie Qi, Hanni Lin, Xiaoli Zhao, and Rui Li contributed equally to this work.

References

1. Bell EF, Hintz SR, Hansen NI, Bann CM, Wyckoff MH, DeMauro SB, et al. Mortality, in-hospital morbidity, care practices, and 2-year outcomes for extremely preterm infants in the US, 2013–2018. *JAMA*. 2022;327(3):248–63.
2. Norman M, Hallberg B, Abrahamsson T, Björklund LJ, Domellöf M, Farooqi A, et al. Association between year of birth and 1-year survival among extremely preterm infants in Sweden during 2004–2007 and 2014–2016. *JAMA*. 2019;321(12):1188–99.
3. Manuck TA, Rice MM, Bailit JL, Grobman WA, Reddy UM, Wapner RJ, et al. Preterm neonatal morbidity and mortality by gestational age: a contemporary cohort. *Am J Obstet Gynecol*. 2016;215(1):e1031–10314.
4. Hamrick SEG, Sallmon H, Rose AT, Porras D, Shelton EL, Reese J, et al. Patent ductus arteriosus of the preterm infant. *Pediatrics*. 2020;146(5):e20201209.
5. Su BH, Lin HY, Chiu HY, Tsai ML, Chen YT, Lu IC. Therapeutic strategy of patent ductus arteriosus in extremely preterm infants. *Pediatr Neonatol*. 2020;61(2):133–41.
6. Clyman RI, Liebowitz M, Kaempf J, Erdeve O, Bulbul A, Håkansson S, et al. PDA-TOLERATE trial: an exploratory randomized controlled trial of treatment of moderate-to-large patent ductus arteriosus at 1 week of age. *J Pediatr*. 2019;205:41–e86.
7. Mirza H, Garcia J, McKinley G, Hubbard L, Sensing W, Schneider J, et al. Duration of significant patent ductus arteriosus and bronchopulmonary dysplasia in extremely preterm infants. *J Perinatol*. 2019;39(12):1648–55.
8. van Laere D, van Overmeire B, Gupta S, El-Khuffash A, Savoia M, McNamara PJ, et al. Application of NPE in the assessment of a patent ductus arteriosus. *Pediatr Res*. 2018;84(Suppl 1):46–56.
9. Ruoss JL, Bazacliu C, Giesinger RE, McNamara PJ. Patent ductus arteriosus and cerebral, cardiac, and gut hemodynamics in premature neonates. *Semin Fetal Neonatal Med*. 2020;25(5):101120.
10. Cervera SB, Saeed S, Luu TM, Gorgos A, Beltempo M, Claveau M, et al. Evaluation of the association between patent ductus arteriosus approach and neurodevelopment in extremely preterm infants. *J Perinatol*. 2024;44(3):388–95.

11. Clyman RI, Hills NK. Prophylactic indomethacin and the risk of serious pulmonary hemorrhages in preterm infants less than 28 weeks' gestation. *J Perinatol.* 2024;44(10):1470–7.
12. Jue JH, Shin SY, Park JH, Kim CS, Choi HJ. Perinatal risk factors for hemodynamically significant patent ductus arteriosus in very low birth weight infants. *Clin Exp Pediatr.* 2026;69(4):313–21.
13. Liu C, Zhu X, Li D, Shi Y. Related factors of patent ductus arteriosus in preterm infants: a systematic review and meta-analysis. *Front Pediatr.* 2021;8:605879.
14. Al-Matary A, Abu Shaheen A, Abozaid S. Use of prophylactic indomethacin in preterm infants: a systematic review and meta-analysis. *Front Pediatr.* 2022;10:760029.
15. de Klerk JCA, Engbers AGJ, van Beek F, Flint RB, Reiss IKM, Völler S, et al. Spontaneous closure of the ductus arteriosus in preterm infants: a systematic review. *Front Pediatr.* 2020;8:541.
16. Semberova J, Sirc J, Miletin J, Kucera J, Berka I, Sebkova S, et al. Spontaneous closure of patent ductus arteriosus in infants ≤ 1500 g. *Pediatrics.* 2017;140(2):e20164258.
17. Hobson AA, Dagle D, Rusin CG, Rhee CJ, Acosta SI, McNamara PJ, et al. Developing a multivariable prediction model to identify a high-volume patent ductus arteriosus in neonates born extremely preterm. *J Pediatr.* 2026;291:114965.
18. Mu J, Li SS, Su AL, Han SP, Zhu JG. [Predictive factors for hemodynamically significant patent ductus arteriosus in preterm infants and the construction of a nomogram prediction model]. *Zhongguo Dang Dai Er Ke Za Zhi.* 2025;27(3):279–85. Chinese.
19. Mitra S, de Boode WP, Weisz DE, Shah PS. Interventions for patent ductus arteriosus (PDA) in preterm infants: an overview of Cochrane Systematic Reviews. *Cochrane Database Syst Rev.* 2023;4(4):CD013588.
20. Shah PS, Lui K, Sjörs G, Mirea L, Reichman B, Adams M, et al. Neonatal outcomes of very low birth weight and very preterm neonates: an international comparison. *J Pediatr.* 2016;177:144–e526.
21. Barry OM, Gudauskas TM, Balzer DT, Bocks ML, Boe BA, Callahan R, et al. Safety and short-term outcomes for infants < 2.5 kg undergoing PDA device closure: a C3PO registry study. *Pediatr Cardiol.* 2023;44(6):1406–13.
22. Cao Y, Jiang S, Sun J, Hei M, Wang L, Zhang H, 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.
23. Guo X, Feng J, Zhao X, Ying E, Liu D, Tu H, et al. Online registry of neonatal necrotising enterocolitis in Shenzhen: protocol for a multicentre, prospective, open, observational cohort study. *BMJ Open.* 2024;14(11):e091290.
24. Giesinger RE, Hobson AA, Bischoff AR, Klein JM, McNamara PJ. Impact of early screening echocardiography and targeted PDA treatment on neonatal outcomes in 22–23 and 24–26 week infants. *Semin Perinatol.* 2023;47(2):151721.
25. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453–7.

26. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63.
27. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
28. Edstedt Bonamy AK, Gudmundsdottir A, Maier RF, Toome L, Zeitlin J, Bonet M, et al. Patent ductus arteriosus treatment in very preterm infants: a European population-based cohort study (EPICE) on variation and outcomes. *Neonatology*. 2017;111(4):367–75.
29. Ay O, Gunes S. Machine learning-based prediction of hemodynamically significant patent ductus arteriosus in preterm neonates: a pioneering insight. *Front Pediatr*. 2025;13:1727418.
30. Weisz DE, Mirea L, Rosenberg E, Jang M, Ly L, Church PT, et al. Association of patent ductus arteriosus ligation with death or neurodevelopmental impairment among extremely preterm infants. *JAMA Pediatr*. 2017;171(5):443–9.
31. Weisz DE, McNamara PJ, El-Khuffash A. Cardiac biomarkers and haemodynamically significant patent ductus arteriosus in preterm infants. *Early Hum Dev*. 2017;105:41–7.
32. Piedvache A, van Buuren S, Barros H, Ribeiro AI, Draper E, Zeitlin J, et al. Strategies for assessing the impact of loss to follow-up on estimates of neurodevelopmental impairment in a very preterm cohort at 2 years of age. *BMC Med Res Methodol*. 2021;21(1):118.
33. Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med*. 2021;40(19):4230–51.

Figures

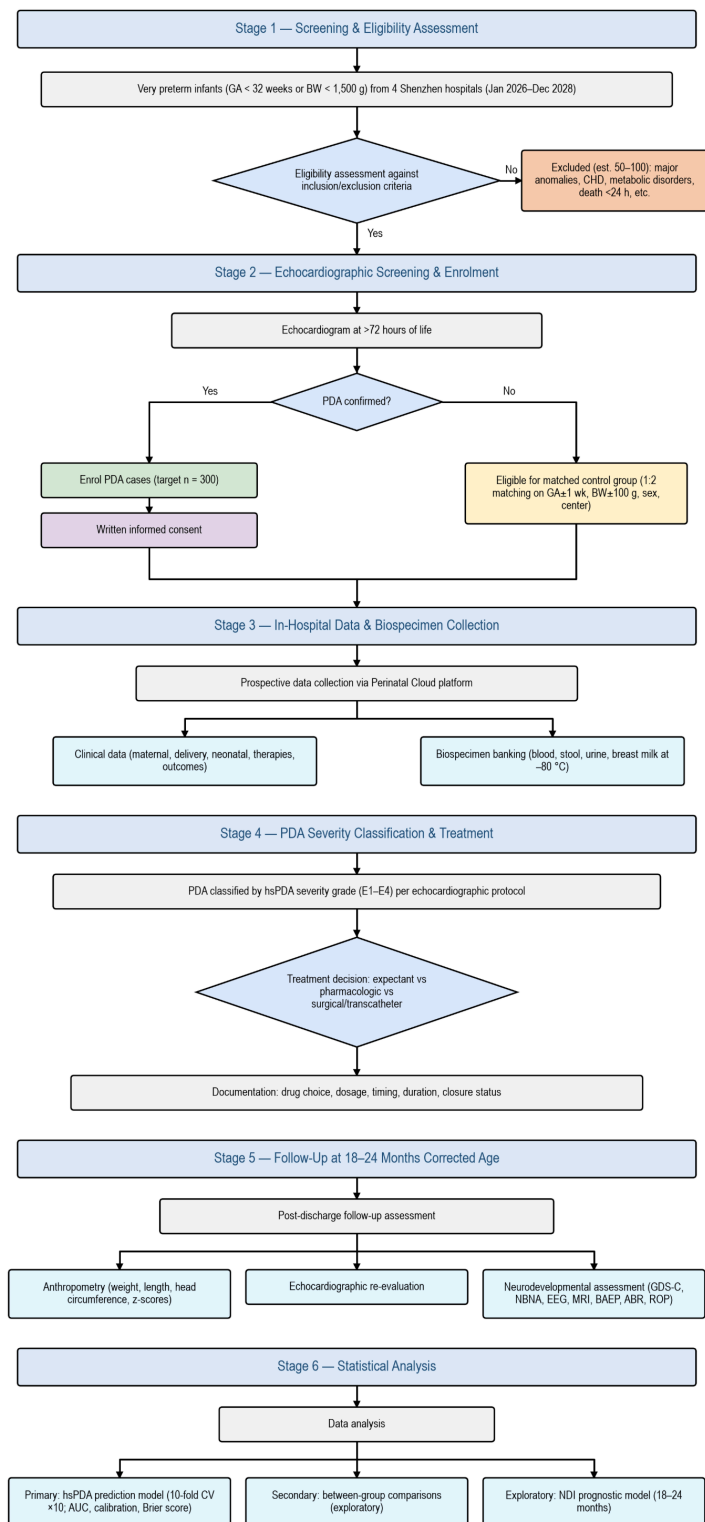


Figure 1

Study flowchart for the Shenzhen Very Preterm Infant PDA Online Registry. Stage 1: Screening of very preterm infants (GA < 32 weeks or BW < 1,500 g) across four centers. Stage 2: Echocardiographic screening at >72 hours; PDA cases enrolled (n = 300) with matched controls (1:2). Stage 3: In-hospital data collection via Perinatal Cloud and biospecimen banking. Stage 4: PDA severity classification (E1–E4) and treatment documentation. Stage 5: Post-discharge follow-up at 18–24 months corrected age.

Stage 6: Statistical analysis including primary hsPDA prediction model development with internal validation (10-fold cross-validation $\times 10$) and exploratory secondary analyses. AUC, area under the receiver operating characteristic curve; BAEP, brainstem auditory evoked potentials; BW, birth weight; CV, cross-validation; GA, gestational age; GDS-C, Griffiths Development Scales-Chinese; hsPDA, haemodynamically significant patent ductus arteriosus; NBNA, Neonatal Behavioral Neurological Assessment; NDI, neurodevelopmental impairment; ROP, retinopathy of prematurity.

Supplementary Files

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

- [PDAProjectPlanRecordForm.docx](#)
- [PDAProjectPlanFlowchart300dpi.tiff](#)