STATISTICAL ANALYSIS PLAN

Management of the Patent Ductus Arteriosus in Premature Infants Trial

(PDA TRIAL) ClinicalTrials.gov number: NCT03456336

SAP VERSION: 1.1

SAP DATE: May 8, 2025

SPONSOR: NICHD

PREPARED BY: RTI International
3040 Cornwallis Rd.
Research Triangle Park, NC 27709-2104

AUTHOR (S): Barbara Do, MSPH
Sonia Thomas, DrPH

SAP HISTORY

Version	Date	Update (topic and section number)	Reviewed by
1.0	Sep 3, 2024	Original	Abhik Das Matthew Laughon
1.1	May 2, 2025	1.Clarified that infant age at inclusion is at time of randomization (4.2) 2.Clarified timing of all assessments in schedule of assessments (4.5) 3.Revised and clarified analysis populations and defined by-treatment and by-medication groups within treatment arms for supplemental displays (Section 5) 4.Added table of pooled centers (7.6) 5.Removed interaction tests with center and race as this was not prespecified in the protocol (7.8) 6.Updated comprehensive list of protocol violations (8.2)	

		7.Removed growth impairment <10% outcomes. Originally included all variations of anthropometric measures in case it was of interest, but growth impairment was not prespecified in the protocol (9.2).	

Contents

LIST OF ABBREVIATIONS

%	percent
~	approximately
<	less than
=	equals
>	greater than
±	plus or minus
≤	less than or equal to
≥	greater than or equal to
α	alpha
AE	adverse event
BPD	bronchopulmonary dysplasia
CA	corrected age
CPAP	continuous positive airway pressure
CV	conventional ventilation
DA	ductus arteriosus
DSMB	data and safety monitoring board
g	grams
GA	gestational age
GDB	Generic Database
GMFCS	Gross Motor Function Classification System
HFV	high frequency ventilation
kg	kilograms
NDI	neurodevelopmental impairment
NEC	necrotizing enterocolitis
NICHD	National Institute of Child Health and Human Development
NIH	National Institutes of Health
NRN	Neonatal Research Network
PDA	patent ductus arteriosus
PMA	post-menstrual age
RCT	randomized, controlled trial
ROP	retinopathy of prematurity
SAE	serious adverse event
SAP	statistical analysis plan
sPDA	symptomatic patent ductus arteriosus

1 BACKGROUND

Patent ductus arteriosus (PDA) is common in premature infants and treatment varies widely. Evidence to provide guidance for treatment of PDA is lacking. During fetal life, the ductus arteriosus (DA) connects the pulmonary artery to the aorta and provides a channel through which the majority of pulmonary arterial blood flow is shunted into the systemic circulation. In most infants, the DA closes shortly after birth. In some infants, especially premature infants, there is a delayed closure of the DA[1], sometimes even as late as after discharge.[2] Approximately 65% of infants born < 28 weeks gestation will have a diagnosis of a PDA at some time during the early neonatal period.[3,4]

In premature infants, PDA is associated with bronchopulmonary dysplasia (BPD). In a large, prospective, population-based study of 1460 infants in North Carolina, PDA was a risk factor for BPD in multivariable analysis, OR=1.9 (95% CI 1.2, 3.1) among infants ventilated at 48 hrs.[5]

The risks and benefits of an active treatment or an expectant management approach need to be evaluated in a randomized, controlled trial (RCT). Because of the uncertainty about which approach is optimal, there is wide practice variation for management of symptomatic patent ductus arteriosus (sPDA). In 2014, 42% of infants <29 weeks gestational age (GA) in the National Institute of Child Health and Human Development (NICHD) Neonatal Research Network (NRN) Generic Database (GDB) who were diagnosed with a PDA were treated with indomethacin or ibuprofen[6] and 18% had their PDA surgically ligated. The percentage of infants with a sPDA who were treated with a COX inhibitor ranged from 19-100% depending on the center; the percentage of infants receiving a PDA ligation ranged from 5-35%. These treatment differences are superimposed on the differences in identification of sPDA which ranged from 14-80% among the centers. Thus, the percent treated for sPDA ranged from 5-60%.

1.1 Protocol History

The first version of the statistical analysis plan (SAP, Version 1.0) was completed based on the version of the protocol dated 19-NOV-2020. If the protocol is updated after version 1 of the SAP, then in any subsequent versions of the SAP, protocol changes that are relevant to the analysis will be summarized here.

Summary of Protocol Changes impacting the SAP	
Date	**Summary of Updates (only including those specific to analysis):**
19-NOV-2020	Protocol version when SAP v 1 was completed

1.2 Changes from planned analyses specified in the Protocol

The protocol (19-NOV-2020 version) specifies that one of the secondary outcomes is the severity of Bronchopulmonary dysplasia at 36 weeks PMA based on NIH consensus definition of moderate or severe. However, since the protocol was first developed, the NIH consensus definition of BPD severity level is no longer used in clinical practice, and instead the Jensen

severity grading is now more commonly used.[7,8] On August 21, 2024 (during study enrollment), the PDA protocol committee unanimously agreed to modify the definition of BPD categories of moderate and severe from the NIH to the Jensen grading criteria. This decision was based on the change in practice over time and involved no evaluation of BPD study data.

Summary of Changes from the Protocol	
Date	**Summary of Change from Protocol to SAP:**
21-AUG-2024 PDA protocol committee meeting	Definition of BPD severity categories updated from the NIH consensus criteria to the Jensen criteria
SAP V1 30-AUG-2024 and SAP V 1.1 1-MAY-2025	The protocol specified that the secondary outcomes of NEC and ROP would be evaluated at 36 months PDA and 22-26 months corrected age. However, this was an inadvertent oversight in the protocol. ROP is collected in the NRN Generic Database (GDB) only at GDB status (defined as the earlier of death, discharge, transfer or 120 days of age), but not at 36 weeks PMA or 2 year follow-up. NEC is recorded in the GDB, but the date of NEC diagnosis is provided, and so NEC will be evaluated based on 36 weeks PMA as specified in the protocol. The SAP is written with the correct timepoint available for ROP.

2 PURPOSE OF THE ANALYSIS

This statistical analysis plan (SAP) contains detailed information about statistical analyses to be performed to conduct statistical interim monitoring and assess the efficacy and safety of active treatment versus expectant management of a sPDA in premature infants. Analyses other than those mentioned in this document can be performed upon further examination of the data and in response to queries from stakeholders and journal reviewers. Such analyses will be clearly labeled as post hoc in any reports. The final results of these analyses will be published in peer reviewed journals and presented at scientific conferences.

3 STUDY OBJECTIVES AND OUTCOMES

3.1 Study Objectives

3.1.1 Primary Objective

- To estimate the risks and benefits of active treatment versus expectant management of a sPDA in preterm infants.

3.1.2 Secondary Objectives

- To determine the incidence of mortality, BPD, and other standard morbidity outcomes of extreme prematurity collected as part of GDB at 36 weeks post-menstrual age (PMA) or at GDB status (defined as the earlier of death, discharge, transfer or 120 days of age).
- To assess the receipt of therapies designed to close the PDA.
- To evaluate anthropometric measures and neurodevelopmental impairment (NDI) at 22-26 months corrected age (CA).

3.2 Outcomes

All outcomes referencing death are with respect to all-cause mortality (i.e., death from any cause), unless otherwise specified. Additional detail regarding the definition of each of the following outcomes can be found in Sections 9 and 10.

3.2.1 Primary Outcome

The primary endpoint of this study is physiologic BPD at 36 weeks PMA, or death occurring prior to 36 weeks PMA.

The physiologic definition of BPD applies to infants who are born at <32 weeks GA (applies to all infants in this trial) and who survive to 36 weeks PMA or are transferred or discharged prior to 36 weeks. The physiologic definition of BPD differs from the traditional definition in two main ways:

First, infants receiving support via ventilator or continuous positive airway pressure (CPAP) at 36 weeks PMA are considered to have BPD by the physiologic definition regardless of whether they are receiving supplemental oxygen or room air.

Second, infants receiving low levels of supplemental oxygen (≤30%) at 36 weeks PMA may be eligible for a physiologic challenge in which there is an attempt to wean the infant to room air. Specifically, infants are eligible for the challenge if at 36 weeks PMA they are receiving effective oxygen <27% and have majority saturation ≥90%, or they are receiving effective oxygen 27-30% and have majority saturation ≥96%, or they are receiving room air by nasal cannula. The challenge takes place between 36 and 37 weeks PMA.

Physiologic PBD determination is summarized as follows:

- Infants receiving supplemental oxygen by positive pressure support via CPAP or ventilator at 36 weeks are considered to have BPD.
- Infants who are successfully weaned to room air during the challenge do not have BPD by the physiologic definition.
- Infants who are on room air before the challenge can take place do not have BPD.

- Infants who are not challenged because their level of support increases (support with CPAP or vent or increased oxygen) are considered to have BPD.
- Infants who fail the challenge are considered to have BPD.
- Infants who are eligible for challenge but who are not challenged because of instability (including surgery or sepsis), or other reasons (such as personnel not available) are classified based on their level of support at 36 weeks.
- Infants who are transferred or discharged before 36 weeks are classified based on the support they are receiving at that time. An infant who is transferred or discharged on supplemental oxygen, ventilator or CPAP at ≤37 weeks PMA will be considered to have BPD (this is used to classify very few cases, if any)

Infants receiving room air by nasal cannula at 36 weeks or at discharge or transfer will have a missing outcome for BPD. This is because the Network determination of whether receipt of room air via nasal cannula constitutes "respiratory support" depends on the flow of air through the cannula. A flow of >.5 liters per minute (lpm) is support, and a flow of ≤.5 lpm is not support. Thus, if we knew the flow rate, we would classify infants receiving room air by nasal cannula with flow >.5 lpm as having BPD by the physiologic definition of BPD and those with flow ≤.5 lpm as not having BPD (this determination was made by the SUPPORT subcommittee). However, the flow rate of air through the cannula is not recorded in the GDB and therefore a determination is not available for these infants. These cases are expected to be < 5% of study infants.

3.2.2 Secondary Outcomes

All of the following are secondary outcomes that will be evaluated, and incidence rates will be reported for each applicable outcome and its corresponding composite with prior death. Outcomes noted with a * are specified in the protocol to be monitored for safety by the DSMB (includes mortality at 36 weeks and GDB status, NEC at 36 weeks PMA, severe ROP at GDB status, as well as mortality and NDI at 2 years corrected age). The DSMB may request other outcomes at their prerogative.

- Mortality: assessed at 36 weeks PMA (component of the primary outcome) and at GDB status (defined as the earlier of death, discharge, transfer or 120 days of age).*
- Bronchopulmonary dysplasia: physiologic test of oxygen therapy at 36 weeks PMA (component of the primary outcome)
- Bronchopulmonary dysplasia severity (moderate, severe): Jensen grading system (grades 2 and 3) at 36 weeks PMA. Note: as described in Section 1.2, the PDA protocol committee modified the definition of moderate and severe BPD from the older NIH consensus grading system to the more commonly used Jensen grading system while the study was ongoing.
- Other standard morbidity outcomes of extreme prematurity collected as part of GDB, including:
    - Necrotizing enterocolitis (NEC) at 36 weeks PMA (NRN definition: proven NEC, no surgery, Stages IIA, IIB, or IIIA AND proven, surgery, Stage IIIB) *

  - Severe retinopathy of prematurity (ROP) at GDB status (NRN definition: Stage 3 or worse in either eye AND as any intervention therapy—retinal ablation, scleral buckle/vitrectomy, Avastin or other anti-VEGF drug) *
- Receipt of therapies designed to close the PDA (NRN definition: ligation or cardiac catheterization) at GDB status and at 2 year follow-up (22-26 months CA):
- Growth at 36 weeks PMA and 2-year follow-up (22-26 months CA): weight, length, and head circumference
- Neurodevelopmental impairment (NDI) at 2 year follow-up (22-26 months CA): NRN definition of moderate NDI (Bayley IV cognitive 70-84, motor 70-84, Gross Motor Function Classification System (GMFCS) 2-3) or severe NDI (Bayley IV cognitive <70, motor <70, GMFCS 4-5, bilateral blind, bilateral hearing loss) or profound NDI (Bayley IV cognitive ≤54, motor ≤46, GMFCS 4-5, bilateral blind, bilateral hearing loss). *
- Mortality at 2-year follow up (22-26 months CA) *

4 STUDY METHODS

4.1 Overall Study Design and Plan

This is a pragmatic randomized multicenter, effectiveness study comparing active treatment of a sPDA to expectant management in premature infants. sPDA and cardiopulmonary compromise will require both clinical and echocardiographic evidence.

Participants with a sPDA allocated to the active treatment arm will receive indomethacin, ibuprofen, or acetaminophen (depending on center preference) within 48 hours of randomization. The choice will be left to the center; however, infants may only receive one medication until discharge, transfer, 120 days in hospital, or death. If the infant receives more than one medication during their stay in the hospital (up to 120 days), it will be considered a protocol deviation. The decision to ligate will be left to the clinical team.

Participants with a sPDA allocated to the expectant management arm will receive supportive care at the clinical team's discretion. Infants assigned to the expectant management arm will receive treatment per local site if cardiopulmonary compromise occurs; treatment may include indomethacin, ibuprofen, acetaminophen, cardiac catheterization, ligation, or continued expected management. Similar to the active treatment arm, receiving more than one medication during their stay in the hospital will result in a protocol deviation.

4.2 Study Population

The study population is defined by the following eligibility criteria.

Inclusion Criteria

1. Postnatal age 48 hours - 21 days at randomization
2. Infant 22 0/7 to 28 6/7 weeks gestation at birth
3. sPDA

Exclusion Criteria

1. Cardiopulmonary compromise (at time of randomization)
2. Known congenital heart disease (besides atrial septal or ventricular septal defect)
3. Known pulmonary malformation (e.g. congenital lobar emphysema, congenital pulmonary adenomatous malformation)
4. Received prior treatment for sPDA
5. Any condition which, in the opinion of the investigator, would preclude enrollment

4.3 Study Arm Assignment and Randomization

Infant will be randomized to either active treatment or expectant management within 48 hours of identification of a sPDA. If the infant is not randomized within 48 hours of identification of a sPDA it will be considered a protocol deviation (for a short delay) or violation (for a long delay or no treatment). An infant must meet all inclusion and none of exclusion criteria at the time of randomization. After meeting eligibility and obtaining consent, randomization to management arms will be 1:1 and stratified by center and GA (22-25 weeks and 26-28 weeks). Eligible neonates from multiple births enrolled in the study will be randomized individually. Treatment arm description is in Section 4.1.

4.4 Masking and Data Lock

4.4.1 General Masking Procedures

Participant families, site staff, and select DCC personnel will be unmasked to treatment allocation, while the senior study statistician (Sonia Thomas) and NRN DCC PI (Abhik Das) will remain masked but may see DSMB data summaries displayed by masked treatment arms.

4.4.2 Database Lock

The database will be locked and unmasked at the completion of study follow-up and any necessary data cleaning. The database may be locked in stages: (1) all 36 week PMA and in-hospital data up to GDB status, (2) all 22-26 month corrected age data.

4.5. Flow Chart of Study Procedures

PROCEDURE	Baseline	36 weeks PMA	GDB status	22-26 months corrected GA
Informed Consent/Privacy Acknowledgement	X			
Demographics	X			
Pertinent Medical History	X			
Medical Baseline Conditions	X			

Body Weight	X[a]	X	X	X
Echocardiogram report	X	X[b]		
Pharmacologic treatments for sPDA		X	X	
Surgical treatments for sPDA		X	X	X
Assessment of BPD		X		
Assessment of death		X	X	X
Assessment of ROP			X	
Assessment of reportable AEs and SAEs		X		
Assessment of protocol deviations and violations		X	X	
Neurodevelopmental Assessment				X
[a] Within 48 hours after enrollment [b] if obtained per clinical care				

5 ANALYSIS POPULATIONS AND TREATMENT GROUPS

Depending on type, each analysis will be conducted within one or more of the following analysis populations. For the overall study analyses for publication, each population will include all infants that meet the population definition.

5.1 All Randomized Infants Population

Since the primary outcome encompasses both efficacy and safety, all analysis for in-hospital outcomes covered by the primary database lock at 36 weeks and GDB status will be based on the same analysis population, which will be comprised of all infants who were randomized, excluding one infant whose parent withdrew consent for use of any data.

Consistent with the intent-to-treat principle, all infants will be analyzed as part of the study arm to which they were assigned by randomization, regardless of actual therapy they received.

Supportive descriptive analyses will further break down the study arms by actual treatment status and type of medication received (see Sections 5.4 and 5.5).

5.2 Per-Protocol (PP) Population

The per-protocol population includes all infants who received treatment according to randomized assignment and per-protocol (i.e., without any protocol violations, including treatment violations

or violations of entry criteria) through 36 completed weeks PMA, discharge, death, or transfer to another hospital, whichever occurs first.

For per-protocol supportive analyses, infants will be analyzed as part of the study arm to which they were assigned by randomization, and all infants with any protocol violation will be excluded. Infants with a protocol deviation will not be excluded.

Protocol violations are defined in Section 8.2. All protocol violation cases will be evaluated in a blinded fashion prior to database lock.

5.3 Special Cases

a. Infants who are withdrawn for consent of use of any data will be noted on the CONSORT diagram but will be excluded from all analyses.

b. If an infant inadvertently switches treatment assignment, they will be analyzed as randomized in all analyses, will be excluded from the per-protocol population, and will be included in the appropriate treatment column for their randomized arm in supportive tables by treatment grouping (see Section 5.5). If deemed warranted at the time of analysis, a further sensitivity analysis of the primary outcome may be performed by switching these infants from the randomized arm to the switched arm.

c. If an infant is immediately withdrawn and does not receive treatment, they will be analyzed as randomized in all analyses and will also be listed on the CONSORT diagram. These infants will be excluded from the per-protocol population (not treated as randomized) and will be excluded from supplemental supportive tables by treatment grouping (see Section 5.5). If deemed warranted at the time of analysis, a further sensitivity analysis of the primary outcome may be performed by removing these infants from the analysis population.

5.4 Infants withdrawn prior to 36 weeks PMA

Infants can be withdrawn from the study by withdrawal of parental consent, or by the study neonatologist if the infant is deemed too medically unstable or for an “other, specify” reason, which includes if an exclusion criterion is identified. Infants who withdraw from the study are given the option to still have their data recorded, and most cases have full data (see special case a in 5.3 if consent for use of data was withdrawn). Reasons for study discontinuation and PDA treatment status will be presented in a supplemental table.

Active treatment situations:

- Withdrew consent for treatment and was not treated (see special case c in 5.3)
- Withdrew from study after treatment initiation.

Expectant management situations:

In the expectant management arm, infants who are withdrawn from the study might still be treated after discontinuation, yet prior to cardiopulmonary compromise. As this would have been a protocol violation if the infant had remained in the trial, these infants will all either have a protocol violation entered by the site of "treated prior to cardiopulmonary compromise", or if not entered by the site, then based on the protocol violation committee review (see Section 8.2) will have an analysis-coded violation of "treated after discontinuation and prior to cardiopulmonary compromise". These cases are excluded from the per-protocol population

Infants discontinued after identification of an exclusion criterion, but later treated prior to cardiopulmonary compromise will have a violation of "ineligible infant randomized". These cases are excluded from the per-protocol population.

In either arm, early withdrawals without a protocol violation are not excluded from the per-protocol population

5.5 Treatment Groupings Based on Treatment Protocol Violations

Efficacy and safety displays will be presented by treatment grouping based on the randomized treatment assignment and the actual treatment administered, as follows:

Active treatment arm:

- A1. Treated within 48 hours of sPDA (as randomized)
- A2. Not treated within 48 hours of sPDA (protocol violation)

Expectant Management Arm:

- E1. Not treated by 36 weeks PMA (as randomized)
- E2. Treated before 36 weeks PMA after cardiopulmonary compromise (as randomized)
- E3. Treated before 36 weeks PMA without or prior to cardiopulmonary compromise (protocol violation)

Note that:

- In the active treatment arm, infants with a protocol deviation of not treated within 48 hours (which represent a minor departure from the protocol) are grouped together with the as-randomized A1 category. This is different from those with a protocol violation which would fall under A2.
- In the active treatment arm, if there are a small number of infants in the A2 category, the A2 group will not be displayed in the tables.

- In the expectant management arm, infants that are treated without or prior to cardiopulmonary compromise, yet the treatment was started after the infant reached 36 weeks PMA are grouped together with the "not treated by 36 weeks PMA" category. However, a supplemental display based on treatments received by GDB status places these infants in the "treated without/before CPC" category

5.6 Treatment Groupings based on medication received by 36 weeks PMA

The 3 medication treatments (indomethacin, ibuprofen, acetaminophen) were selected at the discretion of the treating physician, although it was a protocol violation for infants to receive more than one of these medications through GDB status.

Because there is interest in evaluating some efficacy and safety outcomes by this non-randomized treatment component, some efficacy and safety displays will be presented by medication received within treatment grouping based on the randomized treatment assignment and the actual treatment administered, as follows:

Active treatment 3 categories: indomethacin, ibuprofen, acetaminophen

Expectant management 4 categories: no medication treatment (indomethacin, ibuprofen, acetaminophen) by 36 weeks PMA, indomethacin, ibuprofen, acetaminophen

In these displays:

- Infants in active treatment group who received no treatment are excluded.
- Infants in either arm treated with more than one type of medication are excluded. For all these cases, infants received multiple treatments prior to reaching 36 weeks PMA.
- Infants in the expectant management group who first received medication after 36 weeks are grouped together with the group with no treatment by 36 weeks PMA.
- In the expectant management arm, infants treated before 36 weeks are pooled regardless of initiation after CPC (per protocol) or without/prior to CPC (violation).
- In the expectant management arm, infants treated only with ligation or cardiac catheterization are grouped together with the group with no medication treatment by 36 weeks PMA.

6 SAMPLE SIZE DETERMINATION

Sample size calculations are based on the primary outcome of physiologic BPD or death by 36 weeks' post-menstrual age (PMA) being analyzed prospectively based on the relative risk for the two treatment arms being different by 10%, assuming event rates of 60% vs 50% or 50% vs 40% and that the study will undertake three interim analyses to assess interim efficacy. Power of 80% with overall study-wise two-sided type I error of 0.05 (final analysis type I error 0.044 after multiplicity adjustment for the interim analyses at 25%, 50%, and 75% of total planned enrolled infants using Lan-DeMets α-spending function[9] with an O'Brien-Fleming-type stopping boundary[10]) is achieved with an estimated sample size of 388 per treatment arm, or 776 total.

Therefore, accounting for approximately 5% attrition for the primary outcome at 36 weeks PMA and further adding 2.5% increase for interim looks for a total of 7.5% addition, the **PDA trial aims to enroll a total of 836 infants**, equally allocated to the two treatment arms. Actual attrition rates will be assessed during interim monitoring of the data and sample size recalculation may be considered based on the results, as necessary.

7 STATISTICAL/ANALYTICAL ISSUES

7.1 General Rules

Data will be summarized by treatment group. Categorical measures will be summarized by frequency and percentage; continuous data will be summarized by presenting mean, standard deviation, median, and interquartile range. P-values presented by treatment group will be based on two-sided tests unless otherwise specified, and generally adjusted for randomization stratification factors (center and GA (22-25 weeks and 26-28 weeks). Normality will be checked for continuous outcomes, and if required, transformations or non-parametric methods will be employed.

Statistical computations will be performed, and data summaries created using SAS 9.4 or R.

7.2 Adjustments for Covariates

All analyses by treatment group will be adjusted for randomization stratification factors (NRN center and dichotomous GA (< 26 weeks vs. ≥ 26 weeks). Model-based analyses and test statistics examining the effect of PDA treatment group will be adjusted for NRN study center and GA strata where possible. For example, the primary outcome will be tested via a robust Poisson regression model[9], for which study center and GA strata will be controlled as fixed effects. If enrollment at a specific center is low (less than 10 infants in either GA stratum), then that center may be pooled with another center in close geographic location. See Section 7.6 for pooling of centers. Similar consideration to pooling centers will be given if the primary analysis model faces convergence issues. Alternatively, a random effects model for center may also be used.

Select demographic and baseline (pre randomization) characteristics will be compared between treatment groups. If analyses of these characteristics suggest that substantial differences exist between treatment arms at baseline, their use as explanatory variables will be explored for inclusion as covariates in a secondary supportive model for the primary endpoint to address possible confounding (see Section 9.2).

7.3 Handling of Dropouts and Missing Data

The primary outcome is evaluated at 36 weeks PMA. Most infants in the randomized population will still be receiving inpatient hospital care at 36 weeks PMA and will have a definitive response for the primary outcome (physiologic BPD or death). Missing data for the primary

outcome is expected to be < 5% per treatment arm (based on definition of physiologic BPD, missing physiologic BPD may result from those receiving room air via nasal cannula at discharge or transfer (see Section 3.2.1), and these individuals will not be included in the primary analysis by virtue of the primary analysis method. Should the primary outcome be missing for ≥ 5% per treatment arm, an evaluation will be conducted to determine if there are differences in baseline characteristics for those with and without the primary endpoint (see Section 9.4.3). At database lock, the primary outcome is only missing for 2 (0.4%) infants. Both infants were still in the hospital at 120 days but were missing information needed to determine BPD status.

Secondary efficacy and safety analyses will generally include all available collected data. For these analyses, no data will be excluded, and no imputation will be performed for missing data, unless otherwise specified.

Infants who withdraw from the trial are given the opportunity to consent for continued data collection, allowing for complete treatment and outcome data to be recorded. If the withdrawal impacted the randomized treatment for the infant, a protocol violation was entered.

7.4 Interim Analyses and Data Monitoring

While the study is ongoing, the independent NRN Data Safety and Monitoring Board (DSMB) will examine accumulated data to ensure protection of subjects' safety and assure that the study's scientific goals are being met. Routine monitoring of safety and three formal interim analyses of efficacy are planned for this study. The study subcommittee will review protocol adherence to treatment assignment (i.e., the % crossover) after 50 subjects have been enrolled and reached 36 weeks PMA. The DSMB will then formally review interim safety and efficacy data in a sequential fashion using interim monitoring boundaries after approximately 25% (209), 50% (418), and 75% (627) of the subjects reach the primary outcome at 36 weeks. Treatment groups will be compared statistically using the analysis methods planned for the final analysis (see Section 9.3 for planned efficacy analysis).

The study was stopped by NICHD for futility after the 50% interim analysis.

7.4.1 Interim Analyses for Safety

Formal statistical testing for an imbalance in the composite outcome of mortality and NEC in either treatment arm at GDB status or at 36 weeks PMA, whichever occurs first, will be based on a comparatively liberal Lan-DeMets Pocock boundary[10] at the 3 interim safety looks (25%, 50%, 75%) to guard against any occurrence of false positives while at the same time allowing for stopping at reasonable levels of evidence. Thus, at each interim, an increased incidence of mortality, NEC, or the combination in either treatment group with **$p < 0.0179$** (for 4 total tests at 3 interims plus the final) will be considered as statistically significant evidence of harm that the DSMB can use to recommend suspension of the trial for safety reasons. In addition to the formal

safety outcome, the DSMB will examine other safety outcomes, including all reported serious adverse events by treatment group in considering a recommendation to suspend the trial for safety reasons.

Formal testing of specified safety outcomes at the interim analysis will be conducted on the randomized population, as planned in the final analysis.

Note that although the trial was stopped early and the 75% interim analysis was not completed, the composite of mortality and NEC at 36 weeks PMA will still be evaluated in the final analysis at $p<0.0179$ due to extra unscheduled DMSB meetings conducted during the trial.

7.4.2 Interim Analysis for Efficacy

Three formal interim analyses are planned to evaluate the indication of conclusive evidence of early efficacy in either the active or expectant management treatment group. Interim analyses will be performed when 25%, 50%, and 75% of total planned enrolled infants have been evaluated for the primary outcome (death or physiologic BPD at 36 weeks PMA). To preserve an overall study-wise two-sided Type I error rate of $\alpha = 0.05$, a Lan-DeMets α-spending function[11] with an O'Brien-Fleming-type stopping boundary[12] will be calculated for the efficacy look at the primary outcome, based on 3 planned interim looks. Thus, an increased incidence of the primary outcome (death or BPD at 36 weeks PMA) in either treatment group with $p < 0.000015$ at interim 1 (25%), $p < 0.0030$ at interim 2 (50%), $p < 0.01830$ at interim 3 (75%), and 0.0440 at the final analysis will be considered as statistically significant evidence of efficacy that the DSMB can use to recommend suspension of the trial for efficacy reasons. Exact α levels will be determined based on actual percent of data available at each interim analysis and will be documented in the interim analysis report.

Since the study was stopped after the 50% interim analysis for futility, the 75% interim analysis was not conducted. The final analysis will note that the significance level at study end was to have been 0.044.

The interim analyses will be conducted in the same manner as the primary analysis using robust Poisson regression[7] for the randomized population on the primary outcome, adjusting for GA group and center, to obtain the p-value for comparison of the two treatment groups with the appropriate stopping boundary. However, it is recognized that earlier interim looks may provide too sparse data that may not enable adjusting for center as fixed effect. Center pooling by geographical region will be done in such situations.

7.4.3 Interim Analysis for Futility

Statistical interim futility monitoring will be conducted after 50% and 75% of subjects reach the primary outcome (death or BPD at 36 weeks PMA). At each of these two looks, we will estimate conditional power[12], i.e., the probability to detect a statistically significant treatment benefit of

either treatment group, and its 80% confidence interval given the observed data based on the assumption that the remaining outcome data to be collected will be similar to the data that has been collected thus far. The DSMB may recommend suspension of the trial for futility if this probability is less than **15%**. In addition to the conditional power analysis, the DSMB may also consider other pertinent aspects included in the interim report such as the quality of the data, protocol violations and treatment adherence, rate of enrollment, and rate of attrition in its recommendation to stop for futility.

The conditional power is calculated from the following equation:

$$(CP) = \Phi\left\{\sqrt{\frac{f}{1-f}}Z_1 + \sqrt{\frac{1-f}{f}}(Z_1) - \frac{Z_\alpha}{\sqrt{1-f}}\right\}$$

The confidence interval for conditional power is calculated as:

$$(\mathrm{CP}_L, \mathrm{CP}_U) = \Phi\left\{\sqrt{\frac{f}{1-f}}Z_1 + \sqrt{\frac{1-f}{f}}(Z_1 \pm Z_\gamma) - \frac{Z_\alpha}{\sqrt{1-f}}\right\}$$

- $\Phi\{\cdot\}$ is the cumulative distribution function of the standard normal distribution,
- f is the fraction of information at the time of the interim analysis (percent of planned subjects followed to 36 weeks PMA,
- Z_1 is the interim test statistic (such as $\hat{\theta}_1/\mathrm{SE}(\hat{\theta}_1)$, where $\hat{\theta}_1$ is a consistent estimator of the interim effect size), which is obtained as the Z statistic for the treatment group effect from the robust Poisson regression model. Since either group might be superior, Z will be given a positive value regardless of which group has a lower event rate.
- $Z_{\alpha=0.022} = 2.01$ corresponding to the final one-sided type I error $\alpha = 0.022$, based on the planned interim analysis multiple comparison adjustment.
- $Z_{\gamma=0.10} = 1.28$ for the upper limit of the two-sided 80% confidence interval.

As an example, the following table displays the calculated conditional power and upper confidence limits of conditional power assuming the remainder of the study is the same as the observed data based on the two-sided 80% confidence interval for varying interim test statistics at both the 50% and 75% interim analyses.

	$f = 50\%$		$f = 75\%$	
Interim Test Statistic (Z_1)	Conditional Power (CP)	Upper limit of Conditional Power (CP_U)	Conditional Power (CP)	Upper limit of Conditional Power (CP_U)
0.5	0.032	0.285	0.002	0.016

	$f = 50\%$		$f = 75\%$	
Interim Test Statistic (Z_1)	Conditional Power (CP)	Upper limit of Conditional Power (CP_U)	Conditional Power (CP)	Upper limit of Conditional Power (CP_U)
0.6	0.050	0.357	0.004	0.029
0.7	0.074	0.434	0.008	0.047
0.8	0.106	0.513	0.015	0.075
0.9	0.147	0.592	0.026	0.113
1.0	0.198	0.668	0.043	0.164
1.1	0.258	0.737	0.068	0.227
1.2	0.327	0.798	0.104	0.303
1.3	0.402	0.849	0.152	0.387
1.4	0.481	0.891	0.213	0.478
1.5	0.560	0.924	0.286	0.570
1.6	0.637	0.949	0.370	0.658
1.7	0.709	0.967	0.459	0.738
1.8	0.774	0.979	0.551	0.807
1.9	0.829	0.987	0.640	0.864
2.0	0.875	0.993	0.723	0.908

As a supportive evaluation, the conditional power will also be calculated under the assumption that the data for the remaining subjects in the trial will be as planed in the protocol (specifically, that the planned Z statistic for the treatment effect is 2.01, corresponding to a statistically significant result).

Conditional power under the assumption that the remaining data will be as planned in the protocol is as follows:

$$(CPplanned) = \Phi\left\{\sqrt{\frac{f}{1-f}}Z_1 + \sqrt{\frac{1-f}{f}}\left(Z_{planned}\right) - \frac{Z_\alpha}{\sqrt{1-f}}\right\}$$

And the confidence interval is calculated as:

$$(\mathrm{CP}_L, \mathrm{CP}_U) = \Phi\left\{\sqrt{\frac{f}{1-f}}Z_1 + \sqrt{\frac{1-f}{f}}\left(Z_{planned} \pm Z_\gamma\right) - \frac{Z_\alpha}{\sqrt{1-f}}\right\}$$

7.5 Multicenter Studies

There have been 18 NRN clinical centers eligible to take part in this trial (listed below), with some centers encompassing multiple participating hospitals (sites).

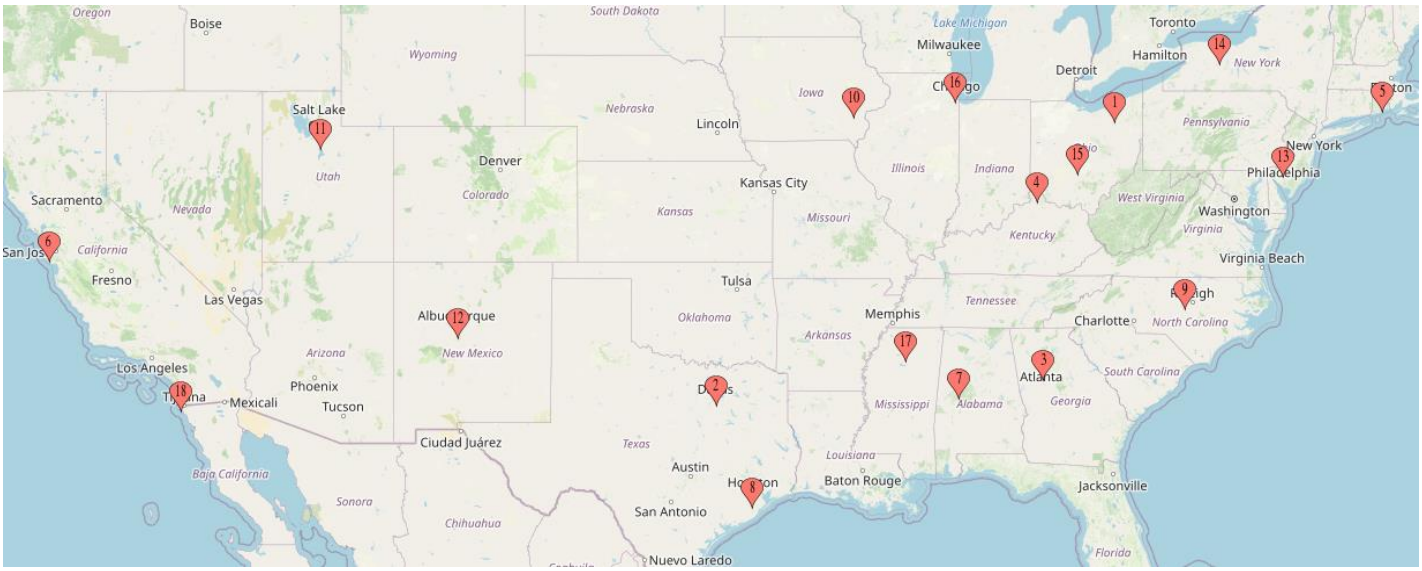
Center ID	Center Name	Center Location
3	Case Western Reserve University	Cleveland, Ohio, USA
4	University of Texas Southwestern Medical Center	Dallas, Texas, USA
9	Emory University	Atlanta, Georgia, USA
11	Cincinnati Children's Hospital Medical Center	Cincinnati, Ohio, USA
14	Women and Infants Hospital of Rhode Island*	Providence, Rhode Island, USA
15	Stanford University	Stanford, California, USA
16	University of Alabama at Birmingham	Birmingham, Alabama, USA
18	University of Texas Health Science Center at Houston	Houston, Texas, USA
19	Duke University	Durham, North Carolina, USA
24	University of Iowa	Iowa City, Iowa, USA
25	University of Utah	Salt Lake City, Utah, USA
26	University of New Mexico Health Science Center	Albuquerque, New Mexico, USA
27	University of Pennsylvania	Philadelphia, Pennsylvania, USA
28	University of Rochester*	Rochester, New York, USA
30	Research Institute at Nationwide Children's Hospital*	Columbus, Ohio, USA
32	Northwestern Medicine†	Chicago, Illinois, USA
33	University of Mississippi Medical Center†	Jackson, Mississippi, USA
34	Sharp HealthCare†	San Diego, California, USA

* Left the NRN March 31, 2023.
† Joined the NRN April 1, 2023.

The number of sites per center is expected to range from one to five and randomization stratification will be based on center and gestational age strata. If models do not converge, we will investigate center pooling based on geography and/or random effects models as alternative strategies. For example, if enrollment of fewer than 10 infants per gestational age stratum occurs within a center, those centers can be pooled with the nearest geographically located center for analysis purposes in order to ensure computational feasibility.

7.6 Pooling Centers

Below is a map of all the centers who participated in this study.



At the end of enrollment, 11 out of 18 centers had <10 infants in at least one GA strata (highlighted in yellow on the following table). Below is a table of final randomization numbers and center pooling.

Combine Map #	Map #	Center	<26 wks GA	≥ 26 weeks GA	Total
1 + 10 + 15 + 16	1	3 = Case Western Reserve University	10	8	18
	2	4 = University of Texas-Dallas	22	14	36
	3	9 = Emory University	12	10	22
	4	11 = Cincinnati Children's	14	16	30
5 + 14	5	14 = Brown University	5	5	10
6 + 12 + 18	6	15 = Stanford University	6	11	17
7 + 17	7	16 = University of Alabama	21	8	29
	8	18 = University of Texas-Houston	41	28	69
	9	19 = Duke University	33	14	47
1 + 10 + 15 + 16	10	24 = University of Iowa	2	2	4
	11	25 = University of Utah	15	30	45
6 + 12 + 18	12	26 = University of New Mexico	4	4	8
	13	27 = University of Pennsylvania	47	27	74
5 + 14	14	28 = University of Rochester	9	10	19
1 + 10 + 15 + 16	15	30 = Ohio State University/Nationwide	12	5	17
1 + 10 + 15 + 16	16	32 = Lurie Children's/Northwestern University	0	2	2
7 + 17	17	33 = University of Mississippi	12	5	17
6 + 12 + 18	18	34 = Sharp Memorial Hospital	8	10	18
		Total	273	209	482

The following table reflects the new pooled centers and updated group totals.

Combine Map #	Combine Center #	Center	<26 weeks GA	≥ 26 weeks GA	Total
1 + 10 + 15 + 16	**3 + 24 + 30 + 32**	**3 = Case Western Reserve University** **24 = University of Iowa** **30 = Ohio State University/Nationwide** **32 = Lurie Children's/Northwestern**	24	17	41
	4	**4 = University of Texas-Dallas**	22	14	36
	9	**9 = Emory University**	12	10	22
	11	**11 = Cincinnati Children's**	14	16	30
5 + 14	**14 + 28**	**14 = Brown University** **28 = University of Rochester**	14	15	29
6 + 12 + 18	**15 + 26 + 34**	**15 = Stanford University** **26 = University of New Mexico** **34 = Sharp Memorial Hospital**	18	25	43
7 + 17	**16 + 33**	**16 = University of Alabama** **33 = University of Mississippi**	33	13	46
	18	**18 = University of Texas-Houston**	41	28	69
	19	**19 = Duke University**	33	14	47
	25	**25 = University of Utah**	15	30	45
	27	**27 = University of Pennsylvania**	47	27	74
		Total	273	209	482

7.7 Multiple Comparisons and Multiplicity

The singular primary outcome will undergo formal hypothesis testing; however, it will be assessed at four different time points (the 3 interim analyses and 1 final analysis). To adjust for multiplicity, the type I error rate will be adjusted using a Lan-DeMets α-spending rule[9] with O'Brien-Fleming-type bounds[10] at each repeated analysis to preserve an overall study-wise α = 0.05 (see Section 7.4.2).

Additionally, formal safety testing of death or NEC will adjust for multiplicity using a Lan-DeMets α-spending function approximating the Pocock boundary[8] (see Section 7.4.1).

All other analyses are considered descriptive or exploratory. Resulting p-values and 95% confidence intervals will generally be provided for descriptive purposes only.

7.8 Examination of Subgroups

Heterogeneity of the treatment effect of the primary outcome variable by center, GA group (as used in randomization) and sex will be investigated by adding suitable interaction terms with treatment group to the robust Poisson model[7] for the primary outcome. Detection of a statistically significant or suggestive interaction (p-value < 0.1) for any of these factors may lead to the relevant subgroup analyses. In particular, the presence of a statistically significant qualitative interaction (which is not expected) would specially require reporting of results within the relevant separate subgroups (for example, if the intervention is shown to be beneficial for infants born ≥26 weeks PMA and harmful for infants born <26 weeks PMA).

Heterogeneity will also be assessed by severity of clinical symptoms, defined based on clinical criteria and PDA size as follows:

- Clinical Criteria
    - Mild: with a small, moderate, or large PDA by echo
    - Moderate: with a small, moderate, or large PDA by echo
    - Severe: with a small or moderate PDA by echo (severe and large is an exclusion criterion)
- PDA size by Echocardiographic evaluation
    - Small: with mild, moderate, or severe clinical criteria
    - Medium: mild, moderate, or severe clinical criteria
    - Large: with mild or moderate clinical criteria (severe and large is an exclusion criterion)

7.9 Assessment Windows

The primary and secondary outcomes assessed at 36 weeks PMA may be collected up to 37 weeks PMA per the study manual of procedures. Long-term secondary outcomes are planned to be assessed at 22 - 26 months CA, however, assessments taking place in a reasonable timeframe outside of this window may be included (e.g., 18 - 30 months CA). Analyses of all primary and secondary outcomes will include all collected data, regardless of the assessment timing.

Additionally, for the primary outcome, the number of assessments obtained outside of the window will be compared between study arms. If there is a difference between study arms, then sensitivity analyses that exclude assessments substantially outside of study window will be conducted to evaluate if results are sensitive to timing of assessments. (see section 9.4.4)

8 STUDY PARTICIPANT CHARACTERIZATION

8.1 Participant Disposition

Participant eligibility status will be summarized, and overall disposition of study participants will be described using a standard CONSORT diagram. The number of participants randomized;

received their randomized intervention; reached NRN GDB status (defined as being discharged, remaining in the hospital at 120 days after birth, dying, or being transferred to another hospital); completed the primary outcome assessment (36 weeks PMA or death); and completed the 2-year follow-up visit will be summarized by study arm.

Additionally, reasons for study withdrawal will be listed, overall and by study arm; reasons for randomized participants who did not receive their randomized intervention will be summarized; and time (days) from birth to NRN GDB status will be summarized by NRN GDB status type.

8.2 Protocol Deviations and Violations

Protocol deviations are identified by site staff, monitors at monitoring visits, and automated checks of the clinical database.

At the completion of the trial and prior to database lock, all protocol deviations and violations will be reviewed by three adjudicators. Adjudicators will be masked to patient and site ID, and to treatment assignment and study outcomes (except where treatment assignment is necessary because it is related to the deviation/violation) and will independently review the list and decide whether the case was a protocol deviation, protocol violation, or a non-reportable event. If additional context is requested before a decision can be made, RTI International will provide that information to all three adjudicators. Once adjudicators finalize their responses, RTI International will compile their answers and flag cases they did not agree on. Adjudicators will then meet to discuss their discrepancies and attempt to meet consensus. If they are unable to meet consensus, they will take the majority vote. All protocol violations will exclude that subject from the per-protocol population. This includes:

- Lack of consent.
- Ineligible infant randomized (did not meet all inclusion/exclusion criteria).
- Not treated as randomized.
    - This includes any case from the active arm who did not receive medication and any case from the expectant management group who was inadvertently treated in the active arm.
- Expectant management arm: received medication, ligation, or cardiac catheterization before cardiopulmonary compromise.
- Active arm: treatment not started within 48 hours of sPDA diagnosis.
- Echocardiogram not performed prior to additional courses of indomethacin, ibuprofen, or acetaminophen.
- Other violations, as identified by PI with violation committee review.

8.3 Study Treatment Exposure and Compliance

The infant will be randomized to either active treatment or expectant management within 48 hours of identification of a sPDA. If the infant is not randomized within 48 hours of identification of a sPDA it will be considered a protocol deviation or violation. An infant must meet all inclusion and none of exclusion criteria at the time of randomization. After meeting

eligibility and obtaining consent, randomization to management arms will be 1:1 and stratified by center and GA (22-25 weeks and 26-28 weeks).

Infants assigned to the active treatment group will receive indomethacin, ibuprofen, or acetaminophen per their local site usual care dosing, formulation (i.e., intravenous or enteral) and schedule within 48 hours of diagnosis of sPDA. The choice will be left to the center; however, infants may only receive one medication during their stay in the hospital (up to 120 days after birth). If the infant receives more than one medication, it will be considered a protocol deviation.

Infants assigned to the expectant management group will receive treatment per local site if cardiopulmonary compromise occurs; treatment may include indomethacin, ibuprofen, acetaminophen, cardiac catheterization, ligation, or continued expected management.

Study treatment exposure will be summarized by study arm, including proportion receiving medication (overall and indomethacin, ibuprofen or acetaminophen), proportion receiving treatment to close the PDA (overall and cardiac catheterization or ligation), and time from sPDA diagnosis to treatment (first treatment with any medication and treatment to close the sPDA).

8.4 Demographic and Baseline Characteristics

The study population will be summarized by study arm for select demographic and baseline characteristics. Summaries will include maternal/household characteristics (e.g., age, race, ethnicity, education, health insurance), pregnancy/delivery characteristics (e.g., use of antenatal steroids, chorioamnionitis, multiple birth, delivery mode), and infant characteristics (e.g., sex, birth weight, small for GA, 1- and 5-minute Apgar scores, use of chest compressions or resuscitation drugs in the delivery room), as well as PDA size (small, medium, large) by electrocardiogram and PDA clinical severity category (mild, moderate, severe). Among the active treatment arm, the receipt of drugs (acetaminophen, ibuprofen, or indomethacin) will also be summarized.

Unadjusted comparisons will be made to identify imbalanced characteristics across study arms. Normally distributed continuous variables will be compared with the Student's t-test, non-normally distributed continuous variables will be compared with the Wilcoxon Rank Sum test, and categorical variables will be compared with a chi-square test or Fisher's exact test for very rare characteristics. P-values from these comparisons will be used for result discussion and covariate selection for the multivariable model described in Section 9.4.2. Only a portion of demographic/baseline characteristics will be considered for model inclusion. See Section 7.2 for additional details on covariates.

8.5 Demographic and Baseline Characteristics of Unenrolled Infants

Demographic and baseline characteristics listed above will be summarized and compared between infants who were enrolled in the study (randomized) versus eligible infants who were not enrolled in the study. This is feasible because information for the non-enrolled infants is available on the NRN GDB database. The goal of this analysis is to determine if the enrolled infants are a representative subset of all eligible infants at these centers.

We will also summarize the rate of treatments to close the PDA (medications, ligation, cardiac catheterization) and mortality at GDB status for unenrolled but eligible infants. We will obtain this information from the NG03 form in GDB which captures whether an infant has PDA, took medication, had ligation, and/or had cardiac catheterization (Y/N format). Note that in the PDA trial, PDA03 captures cases of successful ligations whereas the GDB NG03 form captures whether an infant had attempted ligation, regardless of success. Further, PDA03 notes any doses of acetaminophen, ibuprofen, and indomethacin administered even if it was not specifically given to treat PDA as opposed to GDB NG03 which only captures whether those same medications were given to treat PDA.

9 EFFICACY ANALYSES

9.1 Overview of Efficacy Analysis Methods

All efficacy analyses will be performed using the randomized population, unless otherwise specified.

9.2 Efficacy Variables

The following table identifies the various efficacy variables, defined by response option/interpretation. Clinical explanations of the measures are discussed in Section 3.2. Note that intermittent mandatory ventilation (IMV), conventional ventilation (CV), high frequency ventilation (HFV), and high frequency oscillatory (HFOV) are all considered forms of invasive mechanical ventilation.

Some of the outcomes that come from the GDB are already created in the GDB analysis dataset. In these cases, the variable will be taken from the GDB analysis dataset and not re-created for this trial.

Z-scores will be calculated based on commonly used growth charts for preterm infants.

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
Primary Outcome				
Physiologic bronchopulmonary dysplasia (BPD) or Death by 36 weeks PMA	NG03, NG07 via the GDB analysis dataset: BPD_P, DEATH36	Binary	36 weeks PMA	Physiologic BPD is a standard NRN definition, collected under the Generic Data Base (GDB) protocol. See GDB Manual of Operations, Appendix J for additional details.

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
				Composite measure of all-cause death or positive diagnosis for physiologic BPD at/by 36 weeks PMA. 1 = Yes; Infant died at/before 36 weeks PMA, or was alive and diagnosed with Physiologic BPD at 36 weeks PMA 0 = No; Infant was alive and confirmed negative for physiologic BPD at 36 weeks PMA
Secondary Outcomes				
Death by 36 weeks PMA	NG03 via the GDB analysis dataset: DEATH36	Binary	36 weeks PMA	Infant died at/before 36 weeks PMA. 1 = Yes; Died by 36 weeks PMA 0 = No; Alive at 36 weeks PMA
Physiologic BPD by 36 weeks PMA	NG07 via the GDB analysis dataset: BPD_P	Binary	36 weeks PMA	Positive diagnosis for physiologic BPD at/by 36 weeks PMA. By definition, infants with a non-missing response are alive at the time of the BPD assessment. 1 = Yes; Infant was diagnosed with Physiologic BPD at 36 weeks PMA 0 = No; Infant was confirmed negative for physiologic BPD at 36 weeks PMA

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
Moderate or severe BPD by 36 weeks PMA (Grade 2-3 BPD, by Jensen et al. (2019) [7] definition	NG07 via the GDB analysis dataset: BPD_PRAG4	Binary (for either moderate or severe as well as moderate and severe in 2 categories)	36 weeks PMA	Using the Jensen et al. (2019)[7] definition of BPD severity, infant was diagnosed with Grade 2 or 3 BPD at 36 weeks PMA. 1 = Yes; Grade 2-3 BPD (Moderate or Severe) 0 = No; Grade 0-1 BPD (None or Mild) BPD grade 2: nasal cannula flow >2 lpm, CPAP, or noninvasive ventilation (regardless of FiO2) BPD grade 3: invasive (intubated) respiratory support (regardless of FiO2; includes support via tracheostomy)
Necrotizing enterocolitis (NEC)	NG03	Binary	36 weeks. Note: GDB captures NEC up to status, but dates of NEC confirmed they all occurred prior to 36 weeks and was cross confirmed with the PDA05 AE form for consistency.	Proven NEC 1 = Yes 0 = No
Severe retinopathy of prematurity (ROP)	NG03 via the GDB analysis dataset: ROPSEVERE	Binary	GDB Status	Stage 3 or worse, received intervention/treatment for ROP, or ROP determined severe at status.1 = Yes 0 = No

Death by GDB status	NG02, NG03, NG03E, NG05 via GDB analysis dataset: DEATH	Binary	GDB Status	Infant died at/before GDB status. 1 = Yes; Died by GDB status 0 = No; Alive at GDB status
Receipt of therapies designed to close persistent PDA up to GDB status (ligation or cardiac catheterization)	PDA03	Binary	GDB status Also defined for prior to 36 weeks PMA	Received surgical ligation or cardiac catheterization for PDA closure before GDB status (death, discharge, transfer, or 120 days). 1 = Yes; Infant had surgical ligation and/or catheterization. 0 = No; Infant was confirmed no if responded having neither cardiac catheterization nor surgery.
Receipt of therapies designed to close persistent PDA after discharge (ligation or cardiac catheterization)	PDA08F	Binary	22-26 months CA	Received surgical ligation or cardiac catheterization for PDA closure after GDB status (death, discharge, transfer, or 120 days). 1 = Yes; Infant had surgical ligation and/or catheterization. 0 = No; Infant was confirmed no if responded having neither cardiac catheterization nor surgery.
Neurodevelopmental impairment	NF10A via FU analysis dataset: NDICAT3_4. This variable is 3 levels (mild, moderate, severe). Create binary	Binary	22-26 months CA	NRN definition of moderate NDI (Bayley IV cognitive 70-84, motor 70-84, GMFCS 2-3) or severe NDI (Bayley IV cognitive <70, motor <70, GMFCS 4-5, bilateral blind, bilateral hearing loss) or profound NDI (Bayley IV cognitive ≤54,

	variable (No=1, Yes=2,3)			motor ≤46, GMFCS 4-5, bilateral blind, bilateral hearing loss). 1 = Yes; Moderate or Severe NDI, at least 1 of the criteria was met 0 = No; No or Non-severe NDI, none of the criteria were met
Death by 22-26 months CA	FU analysis dataset: DEATH_FU	Binary	22-26 months CA	Infant died at/before 22-26 months CA. 1 = Yes; Died at/before 22-26 months CA 0 = No; Alive at 22-26 months CA
Growth Outcomes				
Weight at 36 weeks PMA	NG03	Continuous	36 weeks PMA	Weight (g) at 36 weeks PMA
Weight at follow-up	NF05	Continuous	22-26 months CA	Weight (kg) at 22-26 months CA
Length at 36 weeks PMA	NG03	Continuous	36 weeks PMA	Length (cm) at 36 weeks PMA
Length at follow-up	NF05	Continuous	22-26 months CA	Length (cm) at 22-26 months CA
Head circumference at 36 weeks PMA	NG03	Continuous	36 weeks PMA	Head circumference (cm) at 36 weeks PMA
Head circumference at follow-up	NF05	Continuous	22-26 months CA	Head circumference (cm) at 22-26 months CA
Z-score of weight at 36 weeks PMA	NG03	Continuous	At 36 weeks PMA	Z-score of weight at 36 weeks PMA
Z-score of weight at follow-up	NF05	Continuous	At 22-26 months CA	Z-score of weight at 22-26 months CA
Z-score of length at 36 weeks PMA	NG03	Continuous	At 36 weeks PMA	Z-score of length at 36 weeks PMA
Z-score of length at follow-up	NF05	Continuous	At 22-26 months CA	Z-score of length at 22-26 months CA
Z-score of head circumference at 36 weeks PMA	NG03	Continuous	At 36 weeks PMA	Z-score of head circumference at 36 weeks PMA

Z-score of head circumference at follow-up	NF05	Continuous	At 22-26 months CA	Z-score of head circumference at 22-26 months CA
Derived variables				
Received any medication for PDA	PDA03	Binary	GDB status	Received at least 1 medication listed on PDA03 Section A4. Drug Administered 1 = Yes 0 = No
Cardiopulmonary compromise	PDA04	Binary	GDB status	Had large PDA size per local echo report and severe clinical criteria. 1 = Yes 0 = No

9.3 Primary Efficacy Analysis Methods

The primary efficacy outcome of physiologic BPD or death by 36 weeks PMA will be summarized overall and by treatment arm.

The primary efficacy analysis will compare the proportion of infants experiencing the primary outcome for the active arm vs. the expectant arm via a Poisson regression model with model parameter variances estimated by the robust sandwich estimator (robust Poisson regression), by fitting via generalized estimating equation methodology. [9, 14, 15] The primary analysis model will adjust for GA stratum (< 26 weeks vs. ≥ 26 weeks) and NRN pooled study center as fixed effects. The multiplicity-adjusted significance level for each of the interim analyses for the primary outcome are presented in Section 7.4.2, with the final analysis testing against the two-sided significance level of 0.044. A point estimate and 95% confidence interval for the adjusted relative risk will be reported.

The study is powered to find a clinically meaningful risk difference of 10% between the treatment groups for the primary outcome. The protocol development committee felt that a risk difference of 10% is the minimum different that would impact universal acceptance by clinicians to change practices. However, impact on clinical practice is also dependent on neurodevelopmental impairment at 2 years, and a risk difference of 8% may also be clinically relevant. See section 10.4, integration of endpoint and risk for additional outcome interpretation.

9.4 Secondary Analysis Methods for the Primary Efficacy Outcome

Prespecified secondary analyses for the primary outcome of death or physiologic BPD at 36 weeks PMA are outlined in the following subsections. Additional ad hoc analyses motivated by

new information stemming from this clinical trial or other outside research may also be performed.

9.4.1 Per-Protocol and As-Treated Analyses

The primary analysis methods will be replicated for the PP population. These secondary analyses aim to address whether inferences differ for those subjects who did not deviate from the protocol and provide an as-treated assessment of the treatment effect, respectively. Definitions for this population can be found in Section 5.

9.4.2 Adjustment for Baseline Random Imbalance

As a secondary supportive analysis for the primary outcome, additional baseline characteristics may be included as covariates in the model for the primary analysis that were found to be substantially imbalanced across the study arms, despite randomization. See Section 8.4.

If covariates are identified for this supportive model, p-values may be used to identify the impact of the covariates on the outcome. $P< 0.05$ generally determines significance. However, due to the exploratory nature of this analysis, less rigid standards ($p < 0.10$ or $p < 0.20$) may be considered.

9.4.3 Impact of Missing Data

As described in Section 7.3, missing data for the primary endpoint are expected to be minimal (in fact only missing for 2 infants), a supportive analysis comparing baseline characteristics of those with and without the primary endpoint would be conducted to assess the impact of missing data assumptions and handling on the primary study inference. This analysis will be conducted if the amount of missing data for the primary endpoint exceeds 5%.

Due to the minimal amount of missing data expected for the primary outcome, no formal sensitivity analyses of the impact of missing data on the primary analysis will be performed.

9.5 Secondary Efficacy Outcome Analysis Methods

Comparisons of secondary efficacy outcomes between groups will be performed for the ITT population and will be considered descriptive, and not formal tests of hypotheses. All models will adjust for pooled center and GA category as fixed explanatory variables to account for stratification factors.

Incidence of individual components of the primary outcome (death by 36 weeks PMA and physiologic BPD at 36 weeks PMA) and of secondary binary efficacy outcomes (as indicated in the table for efficacy variables in Section 9.2), will be summarized and analyzed in a similar manner as the primary analysis methods described in Section 9.3, i.e., via robust Poisson regression.

Continuous secondary outcomes (weight, length, head circumference) will be analyzed via linear regression with fixed effects for NRN pooled center and GA strata.

Outcomes by drugs (acetaminophen, ibuprofen, or indomethacin) received will also be summarized.

9.6 Exploratory Analysis Methods

New information stemming from this clinical trial or other outside research may motivate additional ad hoc descriptive or hypothesis-generating analyses for both the primary and the secondary efficacy endpoints. Such additional post hoc analyses will be clearly labeled as post hoc in any publication or report.

10 SAFETY ANALYSES

10.1 Overview of Safety Analysis Methods

This section describes safety outcomes that are not also considered efficacy outcomes (described above).

All safety analyses will be performed in an as-treated manner, for descriptive purposes only, and with no special handling for missing data, unless otherwise specified.

Descriptive comparisons of treatment groups will be evaluated via Fisher's exact methods. The p-value for AE deaths will be a supportive p-value to the primary analysis which will be on the primary outcome table.

10.2 Adverse Events (AE)

Reportable AEs will be completed for randomized infants from the time of randomization to 36 completed weeks PMA.

Three specific events (NEC, intestinal perforation, renal insufficiency) and any other event that could be classified as a Serious Adverse Event (SAE) or related to study procedures, in addition to any other AE that results in death, are recorded in this trial.

Note that proven NEC at GDB status is also recorded as part of the GDB. All cases of NEC in this study occurred prior to 36 weeks PMA and align with the NEC AEs captured on PDA05.

Adverse events will be summarized by MedDRA system organ class and preferred term. Summaries will be of the number of individuals experiencing events and will be created for all AEs, AEs by severity, and AEs by relationship to treatment. Summaries will be done for the number and percent of subjects per treatment group experiencing an AE. The number of reportable AEs a subject experiences will also be summarized by treatment group.

For these tables and listings, only monitored on-study AEs will be included, delineated in the table below. On-study AEs include events starting on or after randomization and by 36 weeks PMA. If a complete onset date is unknown and it cannot be confirmed that the event occurred during this time period, then the event will be considered an on-study AE.

Note: AE summaries are also produced for interim safety monitoring reports. See Section 7.4.1 for additional details.

Variable	CRF Source	Type	Collection Timepoint	Definition
Safety Outcomes reported as AEs				
NEC	PDA05	Binary	By 36 weeks PMA	At least one AE for NEC 1= Yes 0= No
Intestinal perforation	PDA05	Binary	By 36 weeks PMA	At least one AE for intestinal perforation 1= Yes 0= No
Renal insufficiency	PDA05	Binary	By 36 weeks PMA	At least one AE for renal insufficiency 1= Yes 0= No
Any other AE that could be classified as a Serious Adverse Event (SAE) or related to study procedures, in addition to any other AE that results in death	PDA05	Binary	By 36 weeks PMA	At least one other reportable AE resulting in death or classified as serious or related to study procedures. 1= Yes 0= No

10.3 Deaths and Serious Adverse Events

SAEs are defined as any AE or suspected adverse reaction that, in the view of either the investigator or sponsor, results in any of the following:

a. Death of infant,
b. Is considered life-threatening,
c. Prolonged hospitalization of infant,
d. Other serious important medical events.

SAEs will be listed for DSMB reports. SAEs, treatment-related SAEs, and SAEs with an outcome of death will be summarized in the manner outlined in Section 10.2. Separate listings and tables summarizing deaths occurring after randomization through 36 weeks PMA, including age at death and cause of death (including primary and contributing causes) will also be created.

Note: death is one of the secondary outcomes, and analysis of death is described in Section 9.5 above.

10.4 Integration of primary endpoint and risks

To determine clinical relevance / recommendations, treatment group risk differences and corresponding 95% CI will be descriptively evaluated for the combination of primary outcome at 36 weeks PMA, mortality at 36-week PMA and 2 years and neurodevelopmental impairment at 2 years, as described in Table 4 of the protocol.

In particular, the protocol development committee felt that clinical recommendations shall be based on a combination of the risk difference of death and the risk difference of NDI, which are competing outcomes.

11 REPORTING CONVENTIONS

Unless required otherwise by a journal, the following rules are standard:

- Moment statistics including mean and standard deviation will be reported at 1 more significant digit than the precision of the data.
- Order statistics including median, min, and max will be reported to the same level of precision as the original observations. If any values are calculated out to have more significant digits, then the value should be rounded so that it is the same level of precision as the original data.
- The median will be reported as the average of the two middle numbers if the dataset contains an even number of observations.
- Test statistics including t and z test statistics will be reported to two decimal places.
- P-values will be reported to 3 decimal places if > 0.001. If it is less than 0.001 then report '<0.001'. Report p-values as 0.05 rather than .05.
- No preliminary rounding should be performed, rounding should only occur after analysis. To round, consider digit to right of last significant digit: if < 5 round down, if ≥ 5 round up.

12 CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

See section 1.2.

13 REFERENCES

1. Van Overmeire B, Van de Broek H, Van Laer P, Weyler J, Vanhaesebrouck P. Early versus late indomethacin treatment for patent ductus arteriosus in premature infants with respiratory distress syndrome. J Pediatr. 2001;138(2):205-211.

2. Herrman K, Bose C, Lewis K, Laughon M. Spontaneous closure of the patent ductus arteriosus in very low birth weight infants following discharge from the neonatal unit. Arch Dis Child Fetal Neonatal Ed. 2009;94(1):F48-50.

3. Chen HL, Yang RC, Lee WT, et al. Lung function in very preterm infants with patent ductus arteriosus under conservative management: an observational study. BMC Pediatr. 2015;15:167.

4. Yum SK, Moon CJ, Youn YA, Lee JY, Sung IK. Echocardiographic assessment of patent ductus arteriosus in very low birthweight infants over time: prospective observational study. J Matern Fetal Neonatal Med. 2017:1-12.

5. Marshall DD, Kotelchuck M, Young TE, Bose CL, Kruyer L, O'Shea TM. Risk factors for chronic lung disease in the surfactant era: a North Carolina population-based study of very low birth weight infants. North Carolina Neonatologists Association. Pediatrics. 1999;104(6):1345-1350.

6. Pranata R, Yonas E, Vania R, Prakoso R. The efficacy and safety of oral paracetamol versus oral ibuprofen for patent ductus arteriosus closure in preterm neonates - A systematic review and meta-analysis. Indian Heart J. 2020;72(3):151-159.

7. Jensen EA, Dysart K, Gantz MG, McDonald S, Bamat NA, Keszler M, Kirpalani H, Laughon MM, Poindexter BB, Duncan AF et al. 2019. The diagnosis of bronchopulmonary dysplasia in very preterm infants. An evidence-based approach. Am J Respir Crit Care Med. 200(6):751–759.

8. Kurihara C, Zhang L, Mikhael M. Newer bronchopulmonary dysplasia definitions and prediction of health economics impacts in very preterm infants. Pediatr Pulmonol. 2021 Feb;56(2):409-417.

9. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159:702-706.

10. Pocock SJ. Group sequential methods in the design and analysis of clinical trial. Biometrika. 1977; 64:191-199.

11. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983; 70:659-663.

12. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trial. Biometrics. 1979; 35:549-556.

13. Lachin JM. A review of methods for futility stopping based on conditional power. Stat Med. 2005;24(18):2747-64.

14. McNutt L-A, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol. 2003;157(10):940–943.

15. Chen W, Qian L, Shi J, Franklin M. Comparing performance between log-binomial and robust Poisson regression models for estimating risk ratios under model misspecification. BMC Medical Research Methodology. 2018;18(63):1-12.

14 LIST OF POTENTIAL DISPLAYS

Data displays may be added, deleted, rearranged or the structure may be modified after finalization of the SAP. Such changes require no amendment to the SAP as long as the change does not contradict the text of the SAP.

Tables
Demographic and Baseline Characteristics Primary Outcome Results • Primary randomized population • Supportive PP population By pre-specified subgroups Secondary Outcomes Results • BPD, NEC, ROP, mortality, NDI outcomes • Growth Outcomes Study Medications AEs and Deaths SAEs Comparison of Eligible Infants, Enrolled vs Not Enrolled
Figures
CONSORT Diagram Kaplan-Meier survival curves of time to death
Data Listings for DSMB Reports
Protocol Deviations and Violations Serious Adverse Events Mortality