STATISTICAL ANALYSIS PLAN

IND Number: 142279

Randomized Controlled Trial of Budesonide + Surfactant versus Surfactant Alone in Extremely Preterm Infants
"The Budesonide in Babies (BiB) Trial"

SAP VERSION: **Version 1.3**

SAP DATE: **September 10, 2024**

SPONSOR: **NIH-NICHD**

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

AUTHORS: **Beth Wiener, DrPH**
Kayla Nowak, BS

APPROVALS

The undersigned acknowledge that they have reviewed the Randomized Controlled Trial of Budesonide + Surfactant versus Surfactant Alone in Extremely Preterm Infants ("The Budesonide in Babies (BiB) Trial") Statistical Analysis Plan (SAP) and agree with the information presented within this document. Changes to this SAP will be coordinated with, and approved by, the undersigned, or their designated representatives.

Signature: Kayla Nowak Date: 09/10/2024

Print Name: Kayla Nowak, BS

Title: Research Statistician

Role: SAP Author

Signature:  Date: 09/10/2024

Print Name: Abhik Das, PhD

Title: Senior Research Statistician
RTI Distinguished Fellow

Role: DCC Principal Investigator

Signature:  Date: 9/10/2024

Print Name: Namasivayam Ambalavanan, MD

Title: Professor of Pediatrics
Director, Neonatal-Perinatal Fellowship Program
Director, Division of Neonatal Research
Co-Director, Division of Neonatology, UAB

Role: Study Principal Investigator

REVISION HISTORY

BiB SAP Revision History

Version	Date	Author/Owner	Description of Change
1.0	15MAY2020	Beth Wiener	Initial version
1.1	01NOV2021	Beth Wiener	Addition of Fisher's exact test for 3% interim safety analysis, approval/signature page, and revision history
1.2	13NOV2023	Beth Wiener	Included allowance for generic budesonide. Note on unconsented infant exclusion from analysis populations. Addition of mITT population. Clarification on pooling sites for analysis. Addition of clustered Poisson GEE analysis if robust Poisson regression has convergence issues. Updated center list for NRN 2023. Clarified per protocol dose is within 10% of expected dose.
1.3	10SEP2024	Kayla Nowak	Finalize Severe NDI definition, following NRN switch to Bayley IV. Add NRN's standard adjudication procedures for 2-year neurodevelopmental outcomes to section 7.3. Clarified alpha-spending procedures for the final efficacy p-value when one or more planned interim analyses were not performed (early study termination). Replaced all planned data listings with data summaries (this trial is Clinical Study Report exempt) Add the option to repeat safety analyses on the PP population. List "Any AE" and "Any AE of Interest" in section 10.2 to match section 3.2.4.1 Removed concomitant medications section, not applicable for this protocol.

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
BiB	Budesonide in Babies
BPD	bronchopulmonary dysplasia
BSID IV	Bayley's Scales of Infant Development, 4th edition
BW	birthweight
CMH	Cochran-Mantel-Haenszel
CMV	cytomegalovirus
CV	conventional ventilation
DCC	data coordinating center
DSMC	data and safety monitoring committee
EDC	electronic data capture
FU	Follow-Up
g	grams
GA	gestational age
GDB	Generic Database
GMFCS	Gross Motor Function Classification System
HFV	high frequency ventilation
ITT	intention-to-treat
IVH	intraventricular hemorrhage
kg	kilograms
LTF	lost to follow-up
MAR	missing at random
MCAR	missing completely at random
mITT	modified intention-to-treat
ml	milliliter
MOP	manual of operations
N/A	not applicable
NDI	neurodevelopmental impairment
NEC	necrotizing enterocolitis
NICHD	National Institute of Child Health and Human Development
NIH	National Institute of Health

NRN	Neonatal Research Network
pct.	percentile
PDA	patent ductus arteriosus
PIE	pulmonary interstitial emphysema
PMA	post-menstrual age
PNA	post-natal age
PVL	periventricular leukomalacia
ROP	retinopathy of prematurity
SAE	serious adverse event
SAF	safety population
SAP	statistical analysis plan
SIP	spontaneous intestinal perforation
SUPPORT	The Surfactant Positive Airway Pressure and Pulse Oximetry Trial
USA	United States of America

1 BACKGROUND AND PROTOCOL HISTORY

Preterm birth is one of the leading causes of neonatal death worldwide (0.97 million, ~35% of all neonatal deaths), and approximately 1 in 8 births in the United States are preterm (Preterm Birth: Causes, Consequences, and Prevention, 2007). Bronchopulmonary dysplasia (BPD) is one of the most common morbidities in extremely preterm infants and is a competing outcome with death. BPD is associated with worse cognitive outcomes in school age children and with abnormal pulmonary function in adolescence and adulthood (Short et al., 2003; Halvorsen et al., 2004; Northway et al., 1990). Use of postnatal steroids to reduce BPD/death is generally minimized due to concerns of neurodevelopmental impairment. However, recent preliminary studies have shown evidence that co-administered surfactant and glucocorticoid (budesonide) reduces the risk of BPD/death (Yeh et al., 2008; Yeh et al., 2016; Venkataraman et al., 2017; Hascoet et al., 2018). Additional studies have shown that budesonide does not accumulate in the central nervous system, suggesting no impact on long-term neurodevelopment (Roberts et al., 2016).

To further assess the safety and efficacy of co-administered surfactant and budesonide, the Neonatal Research Network (NRN) will conduct a large, multicenter clinical trial comparing the effects of combined surfactant and budesonide (intervention, 2.5 ml/kg surfactant + 1.0 ml/kg budesonide) versus surfactant alone (active control, 2.5 ml/kg surfactant) in extremely preterm babies (22 0/7 – 28 6/7 gestation age or birthweight <1000g). For consistency, all sites will use Curosurf as the surfactant and Pulmicort or generic budesonide for the budesonide. The primary outcome is a composite measure of physiologic BPD or death by 36 weeks post-menstrual age (PMA), as these are competing outcomes. This will be a masked, randomized trial with 1:1 treatment allocation. All 15 NRN centers plan to participate in the study; NRN study site (hospital/NICU) will be incorporated into the randomization scheme and analyses via stratification and covariate adjustment.

1.1 Protocol History

The first version of the statistical analysis plan (SAP, Version 1.0) was submitted to the FDA with a version of the protocol dated 15-MAY-2020. In subsequent versions of the SAP, protocol changes that are relevant to the analysis will be summarized here. As of Version 1.2 of the SAP (13-NOV-2023) and Version 1.3 of the SAP (10-SEP-2024), there are no protocol changes that impact the prespecified statistical analyses described in this SAP.

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 early intratracheal co-administration of surfactant and budesonide to extremely preterm infants using data from a randomized, masked, active-controlled clinical trial. The 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

The primary objective is to determine the efficacy of early intratracheal administration of a combination of budesonide with surfactant within 50 hours of birth will reduce the incidence of physiologic BPD or death by 36 weeks post-menstrual age compared to the administration of surfactant alone in extremely premature infants.

3.1.2 Secondary Objectives

The secondary objectives are to describe the differences between co-administered surfactant with budesonide, as compared to surfactant alone, on the incidence of:

- Death by 36w PMA
- Physiologic BPD at 36w PMA
- Higher severity BPD as defined by Jensen et al. (2019) at 36w PMA according to treatment with the following support:
  - No BPD, room air;
  - Grade 1, nasal cannula at flow rates ≤2 L/min;
  - Grade 2, nasal cannula at flow rates >2 L/min or noninvasive positive airway pressure;
  - Grade 3, invasive mechanical ventilation.
- Severe (Grade 3) BPD at 36w PMA according to the Jensen et al. (2019) definition
- Use of additional postnatal steroids (i.e., separate from study drug) for treatment of evolving chronic lung disease by 36w PMA
- Severe neurodevelopmental impairment (NDI) assessed at 2 year follow-up (anticipated 22-26 months (mo) corrected age (CA))
- Death by 2 year follow-up (22-26 mo CA)
- Severe NDI or death as assessed at 2 year follow-up (anticipated 22-26 mo CA)

3.2 Outcomes

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

3.2.1 Primary Efficacy Outcome

The primary outcome is physiologic BPD or death by 36 weeks' post-menstrual age (composite) in all randomized extremely preterm infants and will be recorded using existing NRN Generic Database (GDB) criteria.

3.2.2 Secondary Efficacy Outcomes

The following secondary outcomes will be examined for all randomized subjects:

- Death by 36w PMA
- Physiologic BPD at 36w PMA
- BPD severity as defined by Jensen et al. (2019) at 36w PMA according to treatment with the following support:
  - No BPD, room air;
  - Grade 1, nasal cannula at flow rates ≤ 2 L/min;
  - Grade 2, nasal cannula at flow rates > 2 L/min or noninvasive positive airway pressure;
  - Grade 3, invasive mechanical ventilation.
- Grade 3 BPD at 36w PMA according to the Jensen et al. (2019) definition
- Use of additional postnatal steroids for treatment of evolving chronic lung disease (separate from study drug) from 7 days post final dose (PFD) of study drug to 36w PMA
- Severe NDI assessed at 2 year follow-up (anticipated 22-26 months corrected age), defined as the presence of any of the following:
  - Cognitive composite score on the BSID IV < 70,
  - Motor composite score on the BSID IV < 70,
  - GMFCS level 4-5,
  - Severe hearing impairment (no functional hearing despite amplification), or
  - Bilateral severe visual impairment (bilateral blindness despite correction).
- Death by 2 year follow-up (anticipated 22-26 months corrected age)
- Severe NDI or death as assessed at 2 year follow-up (anticipated 22-26 months corrected age)

3.2.3 Exploratory Outcomes

The following exploratory outcomes will be collected for all randomized subjects:

- BPD severity (none/mild/moderate/severe) by the NIH (NICHD/NHLBI June 2000 Workshop) definition at 36w PMA (Jobe & Bancalari, 2001)
- Number of days on invasive mechanical ventilation by postnatal day 28
- Number of days on invasive mechanical ventilation by 36w PMA

- Intubation after the treatment window (50h postnatal age) and within the first 28 postnatal days
- Intubation after the treatment window (50h postnatal age) and within 36w PMA
- Administration of repeat surfactant doses after the treatment window (50h postnatal age) and within 7 days after last dose of study drug
- Presence of reported respiratory outcomes as assessed at 2 year follow-up (anticipated 22-26 months corrected age), specifically:
  - Recurrent wheezing
  - Chronic coughing

3.2.4 Safety Outcomes

3.2.4.1 Adverse Events

Occurrence of adverse events (AEs) and serious adverse events (SAEs) observed during the study safety monitoring period: initiation of study drug to 7 days (168 hours) after discontinuation of study drug (with the exception of spontaneous intestinal perforation and periventricular leukomalacia being monitored to 30 days after discontinuation of study drug). Study specified AEs are:

- Early-onset sepsis (diagnosed within 72 hours of birth)
  - Moderate: Treated with more than 5 days of consecutive antibiotics/antivirals/antifungals, no cardiovascular instability
  - Severe: Treated for systemic symptoms and with more than 5 days with anti-infective agents, and additional symptomatic support needed, responded to treatment
  - Life-threatening: Same as severe but with cardiovascular instability and slow response or no response to treatment.
- Late-onset sepsis (diagnosed more than 72 hours after birth)
  - Moderate: Treated with more than 5 days of consecutive antibiotics/antivirals/antifungals, no cardiovascular instability
  - Severe: Treated for systemic symptoms and with more than 5 days with anti-infective agents, and additional symptomatic support needed, responded to treatment
  - Life-threatening: Same as severe but with cardiovascular instability and slow response or no response to treatment
- Hyperglycemia
  - Moderate: 120-180 mg/dL on two determinations at least 6 hours apart
  - Severe: More than 180 mg/dL on two determinations at least 6 hours apart, or initiation of insulin
  - Life-threatening: More than 600 mg/dL on two determinations at least 4 hours apart

- Hypertension
  - Moderate: Systolic blood pressure (2-3 determinations over at least 12 hours) which is > 95th percentile but without use of antihypertensive medication
  - Severe: Systolic blood pressure (at least 4 determinations over at least 12 hours) which is > 95th percentile or any use of antihypertensive medication
  - Life-threatening: Sustained increase in BP > 95th percentile for at least 12 hours requiring multiple treatment types or prolonging hospitalization
- Hypotension
  - Moderate: Use of a single vasoconstrictor agent or a single fluid bolus (20 ml/kg or less in 4-hour period)
  - Severe: Use of two vasoconstrictor agents or fluid boluses ≥ 40 ml/kg in 4 hours, or single vasoconstrictor and a fluid bolus 20 ml/kg or more in a 4-hour period
  - Life-threatening: Sustained decrease in BP requiring at least two vasoconstrictor agents for 12 hours
- Prolonged hypoxemia with bradycardia
  - Moderate: Not applicable
  - Severe: Heart rate less than 80 beats per minute for at least 30 seconds, in association with a pulse oximetry reading of less than or equal to 80% for at least 30 seconds
  - Life-threatening: Same as severe but with slow response or no response to treatment
- Endotracheal tube blockage
  - Moderate: Endotracheal tube occlusion requiring tube replacement
  - Severe: Endotracheal tube occlusion requiring tube replacement, associated with prolonged hypoxemia and bradycardia (HR < 80 for ≥ 30 seconds, with SpO_2 < 80 for ≥ 30 seconds)
  - Life-threatening: Endotracheal tube occlusion requiring tube replacement and resuscitation with chest compressions
- Pulmonary air leak
  - Moderate: Pneumothorax with increased observation or needle aspiration OR PIE on x-ray
  - Severe: Pneumothorax with chest tube placement or PIE with use of high frequency ventilation
  - Life-threatening: Chest tube placed; on invasive ventilation with O_2 > 80%
- Intracranial hemorrhage (allowing for three day grace period for head ultrasound)
  - Moderate: Unilateral or bilateral blood/echodensity in the ventricle without distension of the ventricle and without a midline shift; if subarachnoid or subdural blood without a midline shift is detected, it will be included in this category

- Severe: Unilateral or bilateral blood/echodensity filling the ventricle(s) resulting in distension without parenchymal bleeding and without a midline shift
- Life-threatening: Parenchymal hemorrhage/echodensity (including cerebellar hemorrhage) and/or any bleed/echodensity associated with midline shift

Note: Intracranial hemorrhage in preterm infants usually occurs in the first 72 hours (90% by 72h), and is usually determined by a routine head ultrasound on DOL 5-7. As some infants may not get their head ultrasounds on a weekend, there is an additional 3 day grace period.

- Periventricular leukomalacia (PVL) (within 30 days of final study drug administration)
  - Moderate: Per radiology report indication of moderate PVL
  - Severe: Per radiology report indication of extensive PVL
  - Life-threatening: N/A

  Note: PVL is usually diagnosed on the head ultrasound at DOL 28.

- Spontaneous intestinal perforation (SIP) (within 30 days of final study drug administration)
  - Moderate: Not applicable
  - Severe: Intestinal perforation/free air without pneumatosis or bowel necrosis; no shock or respiratory deterioration or hypotension and in the absence of a known gastrointestinal malformation or therapeutic misadventure
  - Life-threatening: Intestinal perforation/ free air; systemic symptoms including, but not limited to tachycardia, hypotension, shock, respiratory deterioration and/or placement of drain or lap
- All other reportable AEs within 7 days of final study drug administration
  - Moderate: The adverse event causes the subject discomfort or interrupts the subject's usual function
  - Severe: The adverse event causes considerable interference with the subject's usual function
  - Life-threatening: The adverse event put the subject at substantial risk of dying at the time of the adverse event (e.g., requiring major surgery or respiratory support)
  - Any event leading to death

3.2.4.2 Clinical Outcomes

Common clinical outcomes of prematurity will also be collected through the NRN Generic Database, namely:

- Death in-hospital prior to 120 days chronologic age
- Necrotizing enterocolitis (NEC) of stage ≥ 2 from birth to NRN GDB status

- Death or NEC (stage ≥ 2) from birth to NRN GDB status, composite
- Patent ductus arteriosus (PDA) from birth to NRN GDB status
- Death or PDA from birth to NRN GDB status, composite
- PDA requiring medical therapy from birth to NRN GDB status
- Death or PDA requiring medical therapy from birth to NRN GDB status, composite
- PDA requiring surgery or cardiac catherization from birth to NRN GDB status
- Death or PDA requiring surgery or cardiac catherization from birth to NRN GDB status, composite
- Severe Retinopathy of Prematurity (ROP) from birth to NRN GDB status
- Death or severe ROP from birth to NRN GDB status, composite

3.2.4.3 Growth Outcomes

Impaired growth outcomes based on preterm infant growth charts will be collected at 36 weeks PMA, namely:

- Weight less than the 10th percentile
- Death by 36 weeks PMA or weight less than the 10th percentile, composite
- Weight Z-score
- Length less than the 10th percentile
- Death by 36 weeks PMA or length less than the 10th percentile, composite
- Length Z-score
- Head circumference less than the 10th percentile
- Death by 36 weeks PMA or head circumference less than the 10th percentile, composite
- Head circumference Z-score

Additional details of the safety outcomes definitions are provided in Section 10.2, 10.4, and 10.5.

4 STUDY METHODS

4.1 Overall Study Design and Plan

This is a phase III, multicenter, masked, active-controlled clinical trial with two treatment arms. Infants will be randomized with 1:1 allocation to one of the two treatment arms: intervention (2.5 ml surfactant + 1 ml (0.25 mg) budesonide / kg) or active control (2.5 ml surfactant / kg). Since surfactant may need to be administered within 30 minutes of birth, antenatal consent and randomization will be sought when possible. We plan to enroll 1160 infants to allow for approximately 5% attrition. Randomization will be stratified by NRN center and gestational age.

The first dose of study-drug must be administered within 48+2 hours of birth. A 2 hour grace period is allowed for attaining and administering the study treatment, provided a decision was made to treat (and randomization occurred) ≤ 48 hours of birth. If a second dose of surfactant is deemed clinically necessary and will be administered within 48+2 hours of birth, the infant will receive the same treatment they had been given at the first dose (intervention or active control). If more than two doses are required, or if the second dose will occur beyond 48+2 hours of birth, surfactant alone will be administered, regardless of participant's assigned treatment arm. Infants will be monitored for adverse events for 7 days (168 hours) after their final study drug administration (max 2 doses). Monitoring will continue for specified adverse events (SIP and PVL) for 30 days after their final study-drug administration. The primary outcome will be assessed at 36 weeks PMA, unless a definitive outcome is reached beforehand (e.g., death, severe BPD, or hospital discharge). Follow-up secondary endpoints will be collected at the two-year follow-up visit planned for 22-26 months adjusted age.

4.2 Study Population

4.2.1 Inclusion Criteria

To be enrolled in this study, participants *must meet* all of the following criteria:

- Liveborn,
- 22 $^{0}/_{7}$ – 28 $^{6}/_{7}$ weeks gestation OR 401 to 1000g birth weight (inclusive),
- ≤ 48 hours postnatal age (dose may be administered up to 50 hours), and
- Clinical decision to administer surfactant.

4.2.2 Exclusion Criteria

To be enrolled in this study, participants *must not meet* any of the following criteria:

- Infants determined to be clinically unstable, terminally ill, or infants with a decision to redirect or limit support, as determined by the clinical team,
- Prior use of surfactant before enrollment (first dose of surfactant must be study drug),
- Infant received systemic steroids prior to enrollment,
- Use of indomethacin, either received by the mother within 24 hours prior to delivery, received by the infant prior to enrollment, or intent to administer to the infant for IVH prophylaxis or PDA management from enrollment up to 7 days from final dose of study drug,
- Infants with serious chromosomal abnormalities or major malformations,
- Infants with known congenital infections, included but not limited to confirmed sepsis and congenital cytomegalovirus (CMV),
- Infants with a permanent neuromuscular condition that affects respiration, or
- Enrollment in a conflicting clinical trial.

4.3 Study Arm Assignment and Randomization

Consent can be obtained either before birth of the infant (antenatal consent) or after birth of the infant (postnatal consent), as long as the consent is obtained early enough to assess eligibility and randomize the infant in the trial by 48h of age. Attempts will be made to obtain antenatal consent, but if not possible, postnatal consent will also be sought to maximize enrollment. A +2h grace period will be allotted for attaining and administering the dose (study drug may be administered up to 50h after birth).

Due to the frequent need to administer surfactant to extremely preterm infants in the delivery room, treatment assignments must be made in an expeditious manner in order to enroll these infants in the study, while also maintaining balanced allocation of the infants to the two treatment arms. If these infants are not enrolled because of time constraints in the delivery room, then inferences may not be generalizable to such infants and enrollment may become unacceptably slow.

To address this issue, the BiB trial adapts the step-forward randomization idea of "Use Next" drug kits, which was originally presented by Zhao et al. (2010). Based on a randomization schedule centrally generated by the NRN Data Coordinating Center (DCC) at RTI, drug kits of budesonide respules or sham will be randomized, assembled, masked, and labeled by the Investigational Drug/Research Pharmacy, then stored in a locked location that will allow for rapid access by the respiratory therapist (or other qualified designated person). Drug kits will be added to the queue such that there are at least three sets of kits available in each gestational age stratum at a given time, and the Investigational Drug/Research Pharmacy will be responsible for ordering, stocking, and supplying the respules of budesonide and sham to the NICU in the drug kits. Two drug kits will be prepared for each randomization number in case of the need for a second dose of study drug within the treatment window, thus a set of drug kits is comprised of two drug kits. Curosurf will be available as clinical supply and separate from the drug kits. As dosing for budesonide may require two respules, at least two respules of budesonide (or sham) will be included in each drug kit, for a total of four respules in a set of drug kits (i.e., four respules per randomization ID).

In addition to labeling the drug kits with the required cautionary statement that the drug product is only for investigational purposes, the sequential randomization number will also be labeled on each kit, such that the designated person (e.g., respiratory therapist) will select the next sequential randomization number in the queue. If eligible multiple births are expected, then the corresponding number of kits will be taken from the queue in the randomization number ordering, as multiple gestation births will be randomized independently. Infants will be considered randomized and assigned to the randomization number when the masked drug kit is opened, which will occur after assessment of eligibility and directly before treatment administration. If the drug kit is not opened, then it may be returned to the queue in the locked location and the randomization number is considered unused. The unmasked respiratory therapist (or other qualified designated person) will mix the budesonide or sham with Curosurf and administer it to the infant. The used respules will then be placed back into the masked drug kits with the infant study number. The Investigational Drug/Research Pharmacy will track usage of the Curosurf, budesonide, and sham by evaluating used vials and respules. Other clinical

caregivers will therefore remain masked to treatment group allocation. In sites where it is necessary that the Investigational Drug/Research Pharmacy prepares the study drug, the Investigational Drug/Research Pharmacist will prepare and supply the prepared study drug to the NICU for administration by the respiratory therapist. Further details about the implementation of the randomization strategy can be found in this study's Manual of Operations.

Due to the need for physical queues of randomized study drug for rapid access, 1:1 randomization to the budesonide + surfactant arm or the surfactant alone arm will be stratified by NRN site and gestational age (< 26 $^0/_7$ weeks vs. ≥ 26 $^0/_7$ weeks), noting that an NRN center may have more than one site (hospital/NICU). Infants who are randomized to the < 26 $^0/_7$ strata but advance to the next GA strata prior to the first dose will still be classified as their originally assigned strata for the purpose of this analysis. Infants who are randomized to the ≥ 26 $^0/_7$ strata but advance to 29 $^0/_7$ weeks GA prior to their first dose (and have a birth weight > 1000g) remain in the intention-to-treat (ITT) population, but should not receive treatment as they are no longer eligible for the study.

All randomization will be performed by the Data Coordinating Center (DCC) at RTI, via the block urn design with a stratum-specific maximum tolerable imbalance known only to the DCC. The block urn design was proposed by Zhao and Weng (2011) as a modification of the urn design and alternative to the permuted block design in order to maintain consistent imbalance control under the maximum tolerable imbalance between treatment groups, while also lowering the probabilities of deterministic assignments. By having lower probabilities of deterministic assignments than the stratified permuted block design, the block urn design decreases the probability of correct guessing, and thus inadvertent unmasking of research staff and possible selection bias.

As an overview, λ balls of one color and λ balls of another color are placed into an active urn, where λ is the maximum tolerable imbalance between the arms and each color corresponds to a treatment arm. As a color-coded ball is selected (representing randomization to a treatment arm), the ball is placed into an inactive urn. Once there is a set of two balls, one of each color, in the inactive urn, these balls are returned to the active urn and are available for selection. The probability of assignment to the intervention arm for the i-th subject is represented by P_{iA},

$$P_{iA} = \frac{\lambda + \min\left(N_{i-1,A}, N_{i-1,C}\right) - N_{i-1,A}}{2\lambda + 2\min\left(N_{i-1,A}, N_{i-1,C}\right) - (i-1)}$$

where N_{iA} is the number of infants allocated to the intervention arm among the i subjects and N_{iC} is the number of infants allocated to the control arm among the i subjects. Thus, determination of the probabilities of treatment assignment take a closed form and are straightforward to calculate. Randomization schedules will be generated via the block urn design and supplied to the sites prior to commencing enrollment.

4.4 General Masking Procedures

This study will be masked, and all attempts will be made to maintain masking to the extent possible. Select DCC personnel will be unmasked to treatment allocation, while the senior study

statistician and NRN DCC PI will remain masked. Only research pharmacists and a designated respiratory therapist (or other qualified person) will be unmasked at the enrolling sites. The designated respiratory therapist (or other qualified person) must be unmasked to allow drug mixing at bedside for expeditious study drug administration; however, this person will not be the primary medical provider nor the primary data collector for that infant, and will not be otherwise involved in the research.

Each site's study-certified pharmacist will be responsible for tracking and documentation of study drug kits, as well as reporting directly to the DCC any protocol deviations/violations regarding unmasking on a designated study form.

At the sites with bedside mixing, the mask will be preserved through use of drug kits. Drug kits will be prepared with budesonide or sham by the unmasked pharmacist and will be fully masked from the exterior in order to maintain masking of clinicians, research personnel, and families. The designated respiratory therapist (or other qualified person) will not know the randomized treatment arm prior to opening the drug kit. Once opened, the budesonide should be mixed with the surfactant by the same unmasked person. Mixing and administration should occur reasonably out of sight from masked individuals, and if needed, masking tape may be used to hide the total volume of the syringe used to administer the mixture (budesonide does not change the appearance of the surfactant solution, although the volume is slightly higher when budesonide is added). Before intratracheal instillation, the syringe will be gently mixed (not shaken) by inverting the syringe, and the surfactant or surfactant + budesonide mixture will be administered in a manner similar to that of routine surfactant therapy. Once the kit is spent, all masked contents of the kit will be returned to the kit, which is resealed and returned to pharmacy for end processing. The randomization ID tied to a kit does not include a coded randomization assignment.

Additional logistical discussion of the masking process are discussed in this study's Manual of Procedures. Considerations for preserving the mask in statistical reports are discussed in Section 4.5 of this document.

4.5 Database Lock

At the end of the study, data lock and unmasking will occur in two stages to ensure complete and accurate data in a final locked data set, yet facilitate timely analysis of the primary outcome data.

1. Primary data lock

    i. Purpose: Primary efficacy outcome, early secondary efficacy outcomes, and all safety outcomes through 36 weeks PMA, used for primary scientific publication as well as for regulatory submission

    ii. Data Included:

        a. All baseline data;

b. All study drug administrations (from study start through maximum 50 hours from birth);

c. All reportable AEs;

d. All reportable concurrent drug exposure through 7 days after last dose of study drug; and

e. All hospital course data and GDB data through NRN status (earliest of discharge/transfer, death, hospitalization up to 120 days chronologic age) including clinical outcomes.

2. Long Term Follow up data lock

i. Purpose: Secondary, long term outcome evaluation

ii. Data Included:

a. All data from 1. above (primary data lock) without change and

b. Neurodevelopmental assessment and respiratory survey at planned 22-26 months corrected age.

This sequential data lock does not result in an interim analysis, but rather an analysis of each of the two target result sections [1) primary clinical outcomes, and 2) secondary, long term efficacy outcomes] being started when the data are complete for that section. In general, no summaries or analyses by treatment group will be provided to any study team member for any data prior to the data being locked. Furthermore, no individuals other than the statisticians at Research Triangle Institute (RTI), the DCC for the NRN, will have access to individual treatment assignment until the end of the first stage of data lock.

Specifically, at the time of the first stage of the data lock, the planned analyses for the associated data will be conducted by DCC statisticians and distributed to the study PI, study subcommittee, and NICHD. At this point, these study team members may become unmasked with treatment assignment for individual study subjects. However, the Follow up investigators who conduct neuro-developmental assessments at the two-year follow-up, the clinical staff, and families will remain masked as to study drug assignment of individual subjects throughout this period. Accordingly, while summary results may be distributed to larger audiences at this time (e.g., as part of new drug application [NDA] submission, conference presentations), any summary results that would be accessible to Follow up investigators, clinical staff or families would not include any information that could unmask individual treatment assignment.

After the completion of the final stage of database lock, Follow-up investigators and clinical staff may be unmasked to individual treatment assignment, if requested.

Any deviations from this plan will be discussed in the study report. For example, this report will include details of any emergency unmasking of individual study subjects due to safety concerns

(e.g., a suspected adverse drug reaction). Likewise, if the study is halted early for safety, futility, or efficacy, some aspects of treatment assignment unmasking may also occur in an expedited fashion.

4.6 Study Flow Chart of Assessments and Evaluations

Procedures	Admission – Dosing Window		Inpatient Monitoring Collection Timepoints					Two-Year Follow-up
	≤ 48h PNA	≤ 50h PNA	7 days PFD[1]	28 days PNA	30 days PFD	36 weeks PMA	120 days PNA	22-26 months CA
Informed Consent[2]	X							
Screening	X							
Baseline Collection	X							
Randomization	X							
Study Drug Administration		X						
Concomitant Medication Monitoring		(X)	X					
Continuous AE Monitoring		(X)	X					
Respiratory Monitoring[3]		(X)	X	X		X		
Neurological Monitoring[4]		(X)	(X)	X				
SIP and PVL AE Monitoring		(X)	(X)	(X)	X			
Growth Outcomes						X		
BPD Assessments						X		
GDB Status Outcomes[5]		(X)	(X)	(X)	(X)	(X)	X	
Follow-up Evaluations								X
Long-term Respiratory Outcomes								X

NOTE: In this table 'X' represents the intended collection point(s), and '(X)' represents an optional collection point if the event is observed prior to the intended collection or if GDB status (discharge, transfer, death, or 120 days PNA) is reached prior to the intended collection point.

[1] PFD = Post-Final Dose [of study drug]. At minimum this is equal to PNA; at maximum this is equal to PNA + 2 days.

[2] Informed consent may be obtained antenatally (i.e., ≤ 0h PNA)

[3] Respiratory Monitoring includes metrics for mode of support, intubation, oxygen levels, vital signs, etc.

[4] Neurological Monitoring includes cranial sonograms, specifically reporting evidence of intracranial hemorrhage, periventricular leukomalacia, or ventriculomegaly observed within 28 days PNA.

[5] GDB Status Outcomes that may extend beyond 36 weeks PMA are clinical outcomes for BiB and are only included for discussion purposes.

5 ANALYSIS POPULATIONS

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.

Note, there was an infant inadvertently randomized and treated after consent was declined. This infant is not included in any of the analysis populations.

5.1 Safety (SAF) Population

The safety population will be comprised of all infants who were randomized and received at least one dose of study drug. The safety population will be used for all safety analyses and infants will be grouped by actual treatment received.

5.2 Intention-to-Treat (ITT) Population

The ITT population is the primary population for formal efficacy analyses. This population includes all subjects randomized. For these analyses, subjects will be analyzed as part of the study arm to which they are assigned by randomization, regardless of actual therapy they received.

5.3 Modified Intention-to-Treat (mITT) Population

The population includes all subjects randomized *and receiving at least one dose of study drug*. For these analyses, subjects will be analyzed as part of the study arm to which they are assigned by randomization, regardless of actual therapy they received.

5.4 Per-Protocol (PP) Population

The per-protocol population includes all subjects who received treatment according to randomized assignment and per-protocol (i.e., without any major protocol violations, including the violation of entry criteria) through 36 completed weeks PMA, discharge, death, or transfer to another hospital with study drug discontinuation or hold occurring only as specified in the protocol.

6 SAMPLE SIZE DETERMINATION

The sample size for the BiB study is determined entirely by comparing the surfactant + budesonide arm to the surfactant alone arm for the primary outcome of physiologic BPD or death measured at 36 weeks PMA. Comparisons of secondary outcomes between the study arms will be considered descriptive and not formal hypothesis tests.

In the NRN, roughly 19% of infants in the GDB do not survive to 36 weeks PMA, and among survivors, 42% are diagnosed with physiologic BPD at 36 weeks PMA, for a combined rate of about 53% for the primary endpoint. Among extremely preterm infants in the GDB (401 - 1000g birth weight or 22 – 28 weeks gestational age), these rates are higher: about 24% do not survive to 36 weeks PMA and roughly 45% of survivors are diagnosed with physiologic BPD at 36 weeks PMA, for a combined rate of about 58% in the target population. It is unlikely that we will see the large magnitude of reduction (24% absolute risk) seen in the trial by Yeh et al. (2016) – it is more likely that we will see a reduction of approximately 10% in absolute risk, from 58% to 48%. It is likely that the outcome could be more common, as we are selecting for the subset of infants who require surfactant. As such, the presented power calculations are conservative in that if the actual risk of BPD or death is greater than 58%, the power associated with an absolute reduction of 10% will be greater than the calculated power.

Assuming the primary outcome of physiologic BPD or death will be analyzed prospectively based on the relative risk for the two treatment arms being different from 1, and that the study will undertake three interim analyses to assess interim efficacy as described in Section 7.4.2, the following table displays estimates for a sufficient sample size per treatment arm based on large sample methods comparing two proportions.

	Two-sided type I error (α)	
Power	Overall α = 0.05 (Final α = 0.044)	Overall α = 0.01 (Final α = 0.009)
80%	406	593
90%	539	753
95%	664	900

For a power of 90% 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), the estimated sample size is 539 per treatment arm, or 1078 total. However, simulation studies for the primary analysis method of the robust Poisson model indicate that a sample size of 550 per group (1100 total) is more accurate to attain 90% power for an overall study-wise type I error of 0.05. Therefore, accounting for approximately 5% attrition for the primary outcome at 36 weeks PMA and randomizing multiple gestation infants independently, the BiB trial will aim to enroll a total of 1160 infants, equally allocated to the two treatment arms. Actual attrition rates will be monitored during interim monitoring of the data, and so the sample size may be recalculated based on the results, as necessary.

Due to the nature of stratifying the randomization by NRN site and two-level gestational age with a stratum-specific maximum tolerable imbalance (MTI) known to the DCC, the overall study arm allocation imbalance could be larger than that which would be expected when stratifying by NRN center and gestational age. If the stratum-specific MTI is defined as λ and there are 15 enrolling NRN centers, then stratifying by NRN center and gestational age would result in an overall MTI of $\lambda \times 2 \times 15 = 30\lambda$. However, when replacing NRN center with NRN site and assuming 33 enrolling NRN sites, the overall MTI is $\lambda \times 2 \times 33 = 66\lambda$. It is of note that reaching this MTI would be very unlikely as every stratum would need to be maximally imbalanced and imbalanced in the same direction. Assuming a total sample size of 1100 (after accounting for loss to follow-up), power sensitivity analyses accounting for the possibility of reaching the MTI (i.e., sample sizes of $550 \pm 33\lambda$ for the two study arms for λ known to the DCC) indicate that 90% power is preserved for the primary analysis.

7 STATISTICAL / ANALYTICAL ISSUES

7.1 General Rules

Data will be summarized by treatment group. In tables and listings, categorical measures will be summarized by frequency and percentage; continuous data will be summarized by presenting mean, standard deviation, median and range; and ordinal data will be summarized by only presenting median and range. P-values presented will be based on two-sided tests unless otherwise specified, and generally adjusted for randomization stratification factors. Normality will be checked for continuous outcomes, and if required, transformations or non-parametric methods will be employed.

Statistical computations will be performed using SAS 9.4 or higher. If additional statistical software is required, this will be discussed in the study report.

7.2 Adjustments for Covariates

All analyses will be adjusted for randomization stratification factors (NRN site and dichotomous GA (< 26 weeks vs. ≥ 26 weeks)). Table summaries will be presented overall, and by GA strata. Model-based analyses and test statistics examining the effect of budesonide will be adjusted for NRN study site and gestational age strata where possible. For example, the primary outcome will be tested via a robust Poisson regression model, for which study site and gestational age strata will be controlled as fixed effects. If enrollment at a specific site is low (less than 10 infants in either gestational age stratum), then that site will be pooled with another site belonging to the same NRN center for analysis purposes. Similar consideration to pooling sites will be given if the primary analysis model faces convergence issues. Low enrolling centers may also be pooled, as described in Section 7.5.

Select demographic and baseline 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 model for the primary endpoint to address possible confounding. Examples of

potential confounders include birth weight, sex, use of antenatal steroids, small-for-gestational-age (SGA), maternal education, and maternal health insurance.

7.3 Handling of Dropouts and Missing Data

The primary analysis for publication is evaluated at 36 weeks PMA. Most infants enrolled in the study 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, and these individuals will not be included in the primary analysis by virtue of the primary analysis method. A sensitivity analysis of the primary outcome to the departure from MCAR missing data (to MAR) will be performed and is described in Section 9.4.3.

Secondary efficacy and safety analyses will generally include available data within the assessment windows. For these analyses, no data will be excluded and no imputation will be performed for missing data, unless otherwise specified. For neurodevelopmental outcomes at 22-26 months corrected age, adjudications may be performed by the NRN BiB and Follow up subcommittees, using investigators masked to treatment arm, in order to assign level of NDI for babies assessed at follow up, but with partially missing neurodevelopmental data.

7.4 Interim Analyses and Data Monitoring

While the study is ongoing, the independent NRN Data and Safety Monitoring Committee (DSMC) 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. All interim analyses will be overseen by the DSMC, an independent monitoring body that is not involved with the conduct of the trial, and the only individuals other than the DSMC who will have access to the results of the interim analysis are the DCC study statistician and a second DCC statistician assigned to validate the results. Study investigators will not have any access to interim data. As the NRN is set up with an independent DCC with its own stream of funding, with long-time, well-established procedures for maintaining masking while collecting data for generating the needed trial reports for the DSMC, the potential for unmasking individuals other than the DCC study statisticians and the DSMC members is minimal. The following three sub-sections detail the interim analysis methods for safety, efficacy, and futility to provide the DSMC the necessary information to recommend suspending or stopping study enrollment. Recommendations from the DSMC are addressed to the Director of NICHD, who has the ultimate responsibility to make decisions to alter or halt this NRN study.

7.4.1 Safety

The DSMC will review safety monitoring reports, generated by the DCC, that include tabular summaries of adverse events (AEs) and study data to determine whether there are any safety concerns that may impact continuation of the trial, or evidence that study procedures should be changed or the trial should be halted, only for reasons relating to the safety of the study subjects or inadequate trial performance (e.g., poor recruitment of subjects). Pre-specified formal safety

looks at the interim data will occur after the first 40 patients (~ 3%) enrolled have been evaluated for the primary outcome (death or BPD at 36 weeks PMA) or been discharged or transferred, and subsequently after 25%, 50% and 75% of the enrolled patients have reached the same milestone. The formal interim safety analyses will be performed on the SAF population and will compare the incidence of serious adverse events (SAEs) or death for those infants receiving budesonide + surfactant versus those receiving surfactant alone using robust Poisson regression adjusted for study site and gestational age group as fixed effects. If enrollment at a specific site is low (less than 10 infants in either gestational age stratum), then that site will be pooled with another site belonging to the same NRN center for analysis purposes. Similar consideration to pooling sites will be given if the model faces convergence issues. Low enrolling centers may also be pooled, as described in Section 7.5. Given that there are 15 centers in the NRN, adjustment for site or center in the robust Poisson regression may not be computationally feasible for the first few interim analyses, in which case generalized linear mixed models (GLMM) with site as a random effect or generalized estimating equation (GEE) methods with site as a clustering variable and an unstructured or exchangeable working correlation structure may be explored. If there is not adequate sample size (e.g., at 3% of enrollment) or if convergence or fit issues persist, then treatment groups will be compared via the common odds ratio with associated 95% confidence interval obtained from Mantel-Haenszel or Fisher's exact method.

Due to the repeated testing, multiplicity adjustment to preserve an overall study-wise two-sided type I error rate of $\alpha = 0.05$ for safety, a Lan-DeMets α-spending function with a Pocock-type stopping boundary will be used (Lan & DeMets, 1983). The adjusted significance level will be 0.0033 at the 3% interim analysis, 0.0196 at 25%, 0.0212 at 50%, 0.0214 at 75%. The exact significance level for each interim analysis will depend upon the timing (or more formally the amount of statistical information available at the time) of the analysis and if/when prior analyses have taken place. Thus, at each interim safety look, a p-value less than the corresponding significance level from comparing the incidence of SAEs/death across groups may be used by the DSMC as evidence of disproportionate harm from the study intervention. For the purpose of formal safety monitoring, SAEs are defined from those safety outcomes specifically listed in Section 3.2.4.1 falling under the definition of SAE in Section 10.3.

7.4.2 Efficacy

Three formal interim analyses are planned to evaluate the indication of overwhelming early efficacy. Interim analyses will be performed when 25%, 50%, and 75% of total planned enrolled infants have been evaluated for the primary outcome (death or BPD at 36 weeks PMA) or been discharged or transferred to another facility. To preserve an overall study-wise two-sided type I error rate of $\alpha = 0.05$, a Lan-DeMets α-spending function with an O'Brien-Fleming-type stopping boundary will be used (Lan & DeMets, 1983). The planned adjusted significance levels are 0.000015 at 25%, 0.0030 at 50%, 0.0183 at 75%, and 0.0440 at the final analysis. The exact alpha for each interim analysis will depend upon the timing (or more formally the amount of statistical information available at the time) of the analysis and if/when prior analyses have taken place. Similarly, the exact alpha for the final analysis will depend on if/when the prior analyses have taken place. Alpha spending allocated to interim analyses that were not performed (i.e., in the event of early study stopping) may be recovered for the final analysis.

The interim analyses will be conducted in the same manner as the primary analysis using robust Poisson regression for the ITT population on the primary outcome, adjusting for gestational age group and site, to obtain the p-value for comparison of the two treatment groups with the appropriate stopping boundary. Interim sample size considerations similarly apply as the interim safety analysis, thus pooling of sites, GLMM, and GEE methods accounting for site may be used as described in Section 7.4.1.

7.4.3 Futility

Futility will be assessed when 50% and 75% of the total planned enrolled infants have been evaluated for the primary outcome (death or BPD at 36 weeks PMA) or been discharged or transferred to another facility. To aid in the assessment of efficacy, the DCC will calculate the conditional power and associated two-sided 80% confidence interval for the primary test of treatment effect based on accrued data for the interim report. The confidence interval for conditional power can be calculated from the following equation,

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

where $\Phi\{\cdot\}$ is the cumulative distribution function of the standard normal distribution, f is the fraction of information at the time of the interim analysis, 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), $Z_{\alpha=0.022} = 2.01$ corresponding to the final one-sided type I error $\alpha = 0.022$, and $Z_{\gamma=0.10} = 1.28$ for the upper limit of the two-sided 80% confidence interval. The DSMC may recommend stopping further enrollment due to a signal for futility if the upper limit of the 80% confidence interval for conditional power (CP_U) is less than 0.50 at the 50% interim analysis or is less than 0.30 at the 75% interim analysis. The following table displays the calculated conditional power and upper confidence limit of conditional power based on the two-sided 80% confidence interval for varying interim test statistics at both the 50% and 75% interim analyses. From this table, it can be seen that the criterion of the upper limit being less than 0.50 at the 50% interim and being less than 0.30 at the 75% interim each corresponds to a conditional power of about 0.10.

	$f = 50\%$		$f = 75\%$	
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
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

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 sites (and gestational age strata) due to treatment administration logistics.

While it is expected that many centers will meet or exceed recruiting the expected number of study subjects (77/center), smaller centers may have enrollment of less than 10 subjects. Enrollment of fewer than 10 infants per gestational age stratum is expected to be rare, but if it

occurs, those centers will be pooled with the nearest geographically located center for analysis purposes.

7.6 Multiple Comparisons and Multiplicity

Only 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 with O'Brien-Fleming-type bounds at each repeated analysis to preserve an overall study-wise $\alpha = 0.05$ (see Section 7.4.2). Additionally, formal interim safety monitoring will adjust for multiplicity using a Lan-DeMets α-spending function approximating the Pocock boundary (see Section 7.4.1).

For model building activities, p-values may be used to identify the best fitting model, as well as covariates to be included in the final models. For these model building activities, p-values < 0.05 will generally determine statistical significance. However, due to the exploratory nature of these models, less rigid standards (such as $p < 0.10$ or $p < 0.20$) may be considered for covariate selection and would be described fully in the study report and peer reviewed publications.

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

7.7 Examination of Subgroups

Heterogeneity of the treatment effect by NRN site or center, gestational age group (as used in randomization), race and sex (as required for NIH funded studies) will be investigated by adding suitable interaction terms with treatment group to the robust Poisson model for the primary outcome. If issues of overparameterization or model convergence arise when assessing heterogeneity of sites, then sites may be grouped, such as by enrollment size categories, geographic location, or site-specific procedures allowing study drug to be kept in queue near the unit (vs. administration by pharmacy per specific infant orders). 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 males and harmful for females).

7.8 Assessment Windows

The primary and secondary outcomes assessed at 36 weeks PMA may be collected up to 37 weeks PMA. Long-term secondary outcomes are planned to be assessed at 22 - 26 months corrected age, however, assessments taking place in a reasonable timeframe outside of this window may be included (e.g., 18 - 30 months corrected age). Additionally, the number of assessments obtained outside of the window of the primary outcome will be compared among study arms. If there are differences among study arms, then sensitivity analyses that

include/exclude assessments outside of study window will be conducted to evaluate if any results are sensitive to timing of assessments.

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 study drug; reached NRN status (defined as being discharged, remaining in the hospital at 120 days, dying, or being transferred to another hospital); completed the primary outcome assessment (36 weeks PMA); 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 study drug will be summarized.

8.2 Protocol Deviations

Protocol deviations are identified by site staff, monitors at monitoring visits, and automated checks of the clinical database. Protocol deviations will be summarized overall and by treatment arm as the number of occurrences by deviation type and the number of subjects with a deviation.

8.3 Study Treatment Exposure and Adherence or Compliance

Each participant will receive up to two doses of study drug (intervention or active-control). A decision to treat and randomization must be made within 48 hours of birth, and dosing (first and optional second) must occur within 50 hours of birth. Dosing and/or randomization outside of these windows will be classified as protocol violations, but will still be included in the ITT population.

Based on clinical experience, it is anticipated that 60% of infants will receive their first dose of study drug in the delivery room, 30% will receive surfactant outside of the delivery room, but within the first 24 hours, and 10% will receive surfactant outside of the delivery room, after 24 hours.

Per protocol, the study drug must be administered intratracheally by a medical professional. To monitor exposure and adherence, the following characteristics will be summarized by treatment arm and center within treatment arm.

- Date/time of each study drug dose (presented as hours since birth for summary tables)
- Actual treatment received for each dose (intervention or active-control)
- Reason first dose NOT received
- Administration of either dose NOT per protocol. Note that actual study drug dose within 10% of the expected dose is deemed per protocol.

- Number of study drug doses administered (0, 1, or 2)
- Number of non-study-drug surfactant administrations within the treatment window (within 50 hours of birth)
- Number of non-study-drug surfactant administrations after the treatment window (more than 50 hours from birth)
- Date/time of non-study-drug surfactant administrations (presented as hours since birth for summary tables).

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 gestational age, 1 and 5 minute Apgar scores, use of chest compressions or resuscitation drugs in the delivery room).

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. 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.

9 EFFICACY ANALYSES

9.1 Overview of Efficacy Analysis Methods

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

9.2 Efficacy Variables

The following table identifies the various efficacy variables, defined by response option/interpretation. Clinical explanation 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.

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
Primary Outcome				

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
9.2.1.1 Physiologic bronchopulmonary dysplasia (BPD) or Death	NG03, NG07	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. 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				
9.2.2.1 Death by 36 weeks PMA	NG03	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
9.2.2.2 Physiologic BPD	NG07	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
9.2.2.3 BPD Severity, by Jensen et al. (2019) definition	NG07	Ordinal	36 weeks PMA	BPD severity, by the Jensen et al. (2019) definition. Grade 1 (Mild) BPD has nasal cannula (NC) flow rate ≤ 2L/min, Grade 2 (moderate) BPD has NC flow rates > 2L/min or non-invasive positive airway pressure, Grade 3 (severe) BPD requires invasive mechanical ventilation. See Jensen et al. (2019) or the GDB

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
				Analysis Variable Reference Manual for additional details (definition may be called the "pragmatic" BPD definition). Ordinal classification of BPD severity: 0 = No BPD 1 = Grade 1 BPD (Mild) 2 = Grade 2 BPD (Moderate) 3 = Grade 3 BPD (Severe)
9.2.2.4 Grade 3 BPD, by Jensen et al. (2019) definition	NG07	Binary	36 weeks PMA	Using the Jensen et al. (2019) definition of BPD severity, infant was diagnosed with Grade 3 BPD at 36 weeks PMA. 1 = Yes; Grade 3 BPD (Severe) 0 = No; No or non-severe BPD
9.2.2.5 Use of additional postnatal steroids	BB04	Binary	36 weeks PMA	Aside from study drug intervention, any other use of postnatal steroids to treat evolving BPD/CLD from 7 days PFD through 36 weeks PMA. 1 = Yes 0 = No
9.2.2.6 Severe neurodevelopmental impairment (NDI)	NF10A	Binary	22-26 months CA	Severe NDI is planned to be a composite of five individual neurodevelopmental assessments. A child will be classified as having severe NDI if one or more of the criteria are met: a.) Bayley's Scale for Infant Development, 4th edition (BSID-IV) cognitive score < 70 b.) Bayley's Scale for Infant Development, 4th edition (BSID-IV) motor score < 70 c.) Gross Motor Functional Classification System (GMFCS) level 4-5 d.) Severe hearing impairment: no functional hearing despite amplification

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
				e.) Severe vision impairment: bilateral blindness, <20/200 vision despite correction See the NRN Follow-Up (FU) Manual of Operations for additional discussion of NDI and the various inputs. Child diagnosed with severe NDI at two-year follow-up. By definition, children with a non-missing response are alive at the time of the NDI assessment. 1 = Yes; Severe NDI, at least 1 of the 5 criteria was met 0 = No; No or Non-severe NDI, none of the 5 criteria were met
9.2.2.7 Death by two-year follow-up	NF10	Binary	22-26 months CA	Child died before two-year follow-up assessment planned for 22-26 months CA. 1 = Yes; Died by 22-26 months CA 0 = No; Alive at 22-26 months CA
9.2.2.8 Severe NDI or Death by two-year follow-up	NF10, NF10A	Binary	22-26 months CA	Composite measure of death or severe NDI at/by 22-26 months CA. 1 = Yes; Child died at/before 22-26 months CA, or was diagnosed with severe NDI at 22-26 months CA 0 = No; Alive and confirmed negative for severe NDI 22-26 months CA
Exploratory Outcomes				
9.2.3.1 BPD Severity, by NIH consensus definition	NG07	Ordinal	36 weeks PMA	BPD severity, by the NIH consensus (NICHD/NHLBI June 2000 Workshop) definition. As an overview, infants are classified as No BPD if they have not received 28 days of respiratory support by 28 days of life. Otherwise, BPD is categorized with Mild as breathing room air ($FiO_2 = 0.21$), Moderate requiring $0.21 < FiO_2 < 0.30$, and Severe requiring $0.30 \leq FiO_2$ and/or PPV or NCPAP at

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
				36w PMA. A more detailed definition can be found in Jobe & Bancalari (2001) and the GDB Analysis Variable Reference Manual. 0 = No BPD 1 = Mild BPD 2 = Moderate BPD 3 = Severe BPD
9.2.3.2 Number of days on invasive mechanical ventilation by postnatal day 28	NG07	Count	By 28 days PNA	Total number of days an infant was on invasive mechanical ventilation through 28 days PNA. Invasive mechanical ventilation will be defined as the infant receiving high frequency ventilation (HFV) or conventional ventilation (CV) as their mode of respiratory support on a given day.
9.2.3.3 Number of days on invasive mechanical ventilation by 36 weeks PMA	NG07	Count	36 weeks PMA	Total number of days an infant was on invasive mechanical ventilation through 36 weeks PMA. Invasive mechanical ventilation will be defined as the infant receiving high frequency ventilation (HFV) or conventional ventilation (CV) as their mode of respiratory support on a given day.
9.2.3.4 Intubation after the treatment window, through postnatal day 28	BB03, NG07	Binary	By 28 days PNA	Infant required intubation after 50 hours PNA (tube not withdrawn after 50 hours PNA, or infant required reintubation > 50 hours PNA and by 28 days PNA). 1 = Yes 0 = No
9.2.3.5 Intubation after the treatment window, through 36 weeks PMA	BB03, NG07	Binary	36 weeks PMA	Infant required intubation after 50 hours PNA (tube not withdrawn after 50 hours PNA, or infant required reintubation > 50 hours PNA and by 36 weeks PMA). 1 = Yes 0 = No
9.2.3.6 Received open label surfactant (non-study-drug)	BB04	Binary	By 7 days PNA	Infant received open label surfactant (separate from study drug) after the treatment window of 50 hours PNA

Variable	CRF Source	Type	Planned Collection Timepoint	Definition
				through 7 days after last dose of study drug. 1 = Yes 0 = No
9.2.3.7 Recurrent Wheezing	BB10	Binary	22-26 months CA	Reported occurrences of wheezing more than twice per week during the worst two-week period in the past 6 months, assessed at 22-26 months CA. 1 = Yes; recurrent wheezing 0 = No; no or nonrecurrent wheezing
9.2.3.8 Chronic Coughing	BB10	Binary	22-26 months CA	Reported occurrences of cough without a cold that lasted more than 3 days in the past 6 months, assessed at 22-26 months CA. 1 = Yes; chronic coughing 0 = No; no or nonchronic coughing

9.3 Primary Efficacy Analysis Methods

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

The primary efficacy analysis for the ITT population will compare the proportion of infants experiencing the primary outcome for the surfactant + budesonide arm vs. the surfactant alone 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 (McNutt et al., 2003; Zou, 2004; Chen et al., 2018). The primary analysis model will adjust for gestational age stratum (< 26 weeks vs. ≥ 26 weeks) and NRN study site as fixed effects. More formally, the point and interval estimates and hypothesis tests will be obtained using the following statistical model,

$$\log(\boldsymbol{\mu}) = \boldsymbol{X\beta}$$

where $\log(\cdot)$ denotes the natural logarithm function; $\boldsymbol{\mu} = [\mu_i]' = \left[E[y_i]\right]' = E[\boldsymbol{y}]$ is the vector of expected values for y_i, where y_i equals 0 or 1 corresponding to surviving without BPD to 36 weeks PMA or being diagnosed with BPD or dying by 36 weeks PMA, respectively, for the ith infant; $\boldsymbol{X}$ is the design matrix for the explanatory variables using reference cell coding; $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \beta_2, \beta_{3,1}, \ldots, \beta_{3,k-1}\right)'$ is the vector of model parameters for the explanatory variables encoded in $\boldsymbol{X}$, with β_0 being the intercept (overall reference level) parameter, β_1 the parameter for the surfactant + budesonide effect (vs. surfactant alone), β_2 the parameter for the effect of <

26 weeks gestational age (vs. ≥ 26 weeks), and $\beta_{3,1}, \ldots, \beta_{3,k-1}$ the parameters for the effects of sites 1 through $k-1$ (vs. site k). Solutions to the generalized estimating equations, $\widehat{\boldsymbol{\beta}}$, are consistent estimates of the model parameters $\boldsymbol{\beta}$, for which valid standard errors are obtained using the robust (empirical) sandwich estimator of $\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}\right)$ (Liang & Zeger, 1986).

The primary hypothesis will be tested by incorporating the robust variance estimator for the treatment effect in the robust Poisson model where the null hypothesis of no treatment effect on the probability of BPD or death by 36 weeks PMA (H_0: $\beta_1 = 0$) is rejected in favor of the alternative hypothesis (H_A: $\beta_1 \neq 0$) at one of the interim analyses or the final analysis with a p-value falling below the significance level consistent with an overall study-wise two-sided type I error rate of α = 0.05. 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 at 100% enrollment.

The primary analysis will be implemented following the sample SAS code below for the ITT population (`ITT=1`), where study arm assignment `RNDARM` is 0 if randomized to the surfactant alone arm and 1 if randomized to the surfactant + budesonide arm (corresponding to $\hat{\beta}_1$). The primary outcome is binary `BPD_DEATH`, for which 0 denotes a favorable outcome (i.e., alive without BPD at 36 weeks PMA) and 1 denotes an unfavorable outcome (i.e., diagnosed BPD or death by 36 weeks PMA). `RANDSTRAT` is the dichotomous gestational age randomization strata where 1 denotes gestational age < 26 weeks and 2 denotes gestational age ≥ 26 weeks (corresponding to $\hat{\beta}_2$), `SITE` represents the categorical variable for NRN study site (corresponding to $\hat{\beta}_{3,1}, \ldots, \hat{\beta}_{3,k-1}$), and `SUBJID` is the unique subject ID for each infant. The `EXP` option for the `ESTIMATE` statement provides the estimate and confidence limits for the covariate-adjusted relative risk, `E` prints the contrast matrix coefficients for verification purposes, and `ALPHA=SIG` defines the two-sided significance level `SIG` = α for $(1-\alpha)$ confidence intervals for the covariate-adjusted relative risk, adjusted for multiplicity.

```
PROC GENMOD;
  WHERE ITT = 1;
  CLASS RNDARM(REF="0") RANDSTRAT(REF="2") SITE SUBJID / PARAM=REF;
  MODEL BPD_DEATH = RNDARM SITE RANDSTRAT / DIST=POISSON LINK=LOG;
  REPEATED SUBJECT = SUBJID / TYPE=IND;
  ESTIMATE "S+B vs S for Pr(BPD/Death)" RNDARM 1 / EXP E ALPHA=SIG;
RUN;
```

If enrollment at a specific site is low (less than 10 infants in either gestational age stratum), then that site will be pooled with the next smallest site belonging to the same NRN center for covariate adjustment in the primary analysis model. Similar consideration to pooling sites will be given if the model faces convergence issues. Low enrolling centers may also be pooled, as described in Section 7.5. Furthermore, it is possible that adjustment for site or center in the robust Poisson regression may not be computationally feasible, in which case GLMM with site as a random effect or GEE methods with site as a clustering variable and an unstructured or exchangeable working correlation structure may be used as the primary analysis model.

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 that were used to analyze the ITT population will be replicated for the mITT population, PP population, and SAF population. These secondary analyses aim to address whether inferences differ for those subjects who did not receive study drug, did not deviate from the protocol, and to provide an as-treated assessment of the treatment effect, respectively. Definitions for these populations can be found in Section 5.

9.4.2 Adjustment for Baseline Random Imbalance

As a secondary analysis for the primary outcome, baseline factors known to affect BPD (e.g., birth weight, sex, race/ethnicity) will be included as covariates in the robust Poisson model specified as the primary analysis for the composite outcome of physiologic BPD or death by 36 weeks PMA. In addition to adjusting for site and gestational age strata as fixed effects, additional baseline characteristics may be included as covariates in the model that were found to be imbalanced across the study arms, despite randomization, as described in Section 8.4. This model can be fit using stepwise or backward selection if the number of explanatory variables is too large. A point estimate and 95% confidence interval for the adjusted relative risk will be reported for the final model.

9.4.3 Impact of Missing Data

Although missing data for the primary endpoint are expected to be minimal, a sensitivity analysis will 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%. Multiple imputed datasets will be generated from an imputation model that includes the explanatory variables from the primary analysis, in addition to ancillary variables that are thought to explain the missingness or predict mortality or BPD at 36 weeks PMA. The number of data sets imputed will be equal to 100 times the fraction of missing outcome data (i.e. if 25% of infants are missing the primary outcome, 25 data sets will be imputed), with a minimum of 10 imputed data sets. Each imputed dataset will be analyzed separately using the statistical methods described in Section 9.3, and parameter estimates will be pooled using Rubin's combination rules to produce estimates and standard errors that incorporate missing data uncertainty (Rubin, 1987). This multiple imputation approach assumes values are missing at random (MAR) and provides an evaluation of impact of missing data uncertainty. Therefore, due to the utilization of generalized estimating equation methods for the primary analysis, this sensitivity analysis is namely evaluating the robustness of the primary analysis to departures from the assumption that the primary endpoint is missing completely at random (MCAR). Due to

the minimal amount of missing data expected for the primary outcome, further sensitivity analyses on the impact of missing data will not be performed.

9.5 Secondary Efficacy Outcomes 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 site and gestational age 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 will be analyzed via linear regression with fixed effects for NRN site and gestational age strata.

Ordinal secondary outcomes (such as BPD severity/grade), will be analyzed via proportional odds models with fixed effects for NRN site and gestational age strata. If the proportional odds assumption does not hold, a partial proportional odds model or Mantel-Haenszel mean score test (with scores determined by standardized midranks, a.k.a. modified ridits) will be used to analyze the ordinal data. Nominal outcomes (discrete outcomes with more than two levels without inherent ordering) will be analyzed via generalized logit models. If convergence or fit issues arise with the generalized logit model, then the nominal outcomes may be analyzed via the Mantel-Haenszel general association test (with degrees of freedom equal to the number of outcome categories minus one).

Secondary count outcomes will be analyzed via Poisson regression with fixed effects for NRN site and gestational age strata, accounting for person-time via an offset when applicable for a rate with varying exposure times. In the presence of over- or underdispersion, negative binomial regression or a scaling parameter may be explored to better fit the data. Specifically, if an excessive amount of zeros are observed, a zero-inflated model may be fit, such as a zero-inflated Poisson model or zero-inflated negative binomial model. In contrast, if the observed mean count is large and the count data approximates a normal distribution relatively well, traditional linear regression may be used to model the outcome as continuous.

New information stemming from this clinical trial or other outside research may motivate additional ad hoc descriptive or hypothesis-generating analyses for the secondary efficacy endpoints. Additional post hoc analyses may be performed on exploratory endpoints.

10 SAFETY ANALYSES

10.1 Overview of Safety Analysis Methods

All safety analyses will be performed on the safety (SAF) population in an as-treated manner, for descriptive purposes only, and with no special handling for missing data, unless otherwise

specified. Safety analyses may be repeated on the per protocol (PP) population in an as-treated manner if the PP population excludes SAF participants with violations that are suspected to impact safety outcomes (e.g., eligibility, treatment exposure, concomitant medications). Descriptive comparisons of treatment groups will be provided for most safety table summaries and will be obtained through robust Poisson regression for binary outcomes with fixed effects for NRN site and gestational age strata. Comparisons between treatment groups will be described with estimated relative risks and associated 95% confidence intervals for descriptive purposes only. In the presence of convergence issues for the robust Poisson model, adjustment for NRN site as a random effect will be considered. If convergence or fit issues persist, then treatment groups will be compared via the common odds ratio with associated 95% confidence interval obtained from Mantel-Haenszel methods for cross-classifications of NRN site and gestational age strata that offer adequate cell sample sizes or Fisher's exact methods.

10.2 Adverse Events

Reportable AEs include events starting or worsening in severity after start of study drug through 7 days after last study drug dose. AEs will be reported and graded using the Toxicity Table for Premature Neonates: NICHD NRN (Appendix A of this study's Manual of Operations). Adverse events that are mild in severity, transient, and expected in the study population will not be recorded unless it is considered unexpected for this study. An event that is present at baseline is not an AE unless there is an increase in grade (i.e., worsening in severity or frequency) or the event resolves and then returns. Using the Toxicity Table, events will be summarized by category and preferred event term. Summaries will be of the number of individuals experiencing events (occurring with moderate or higher severity as indicated in Appendix A of this study's Manual of Operations) 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, only monitored on-study AEs (listed in Section 3.2.4.1) will be included, delineated in the table below. On-study AEs include events starting on or after Day 1 and prior to 7 days after last dose, with the exception of SIP and PVL being monitored until 30 days after the last dose. 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				
Any AEs	BB05, NG03	Binary	By 30 days post-treatment	At least one AE reported for the study. 1= Yes 0= No

Variable	CRF Source	Type	Collection Timepoint	Definition
Any AEs of interest	BB05, NG03	Binary	By 30 days post-treatment	At least one pre-specified AE reported for the study (i.e., at least one AE that is not classified as an "Other AE"). 1= Yes 0= No
Early-onset sepsis	BB05, NG03	Binary	≤ 72 hours PNA	At least one AE for sepsis, and reported early-onset sepsis. 1= Yes 0= No
Late-onset sepsis	BB05, NG03	Binary	> 72 hours PNA, by 7 days post-treatment	At least one AE for sepsis, and reported late-onset sepsis. 1= Yes 0= No
Hyperglycemia	BB05	Binary	By 7 days post-treatment	At least one AE for elevated blood glucose. 1= Yes 0= No
Hypertension	BB05	Binary	By 7 days post-treatment	At least one AE for elevated blood pressure. 1= Yes 0= No
Hypotension	BB05	Binary	By 7 days post-treatment	At least one AE for low blood pressure. 1= Yes 0= No
Prolonged hypoxemia and bradycardia	BB05	Binary	By 7 days post-treatment	At least one AE for combined low heart rate and low pulse oximetry reading for at least 30 seconds. 1= Yes 0= No
Endotracheal Tube Blockage	BB05	Binary	By 7 days post-treatment	At least one AE for endotracheal tube blockage. 1= Yes 0= No
Pulmonary Air Leak	BB05	Binary	By 7 days post-treatment	At least one AE for pulmonary air leak. 1= Yes 0= No
Intracranial Hemorrhage	BB05	Binary	By 7 days post-	At least one AE for intracranial hemorrhage.

Variable	CRF Source	Type	Collection Timepoint	Definition
			treatment (with allowance of 3 day grace period for head ultrasound)	1= Yes 0= No
Any other AE	BB05	Binary	By 7 days post-treatment	At least one other reportable AE, as defined in Appendix A of the MOP. 1= Yes 0= No
Spontaneous Intestinal Perforation (SIP) without NEC	BB05	Binary	By 30 days post-treatment	At least one AE for SIP without NEC. 1= Yes 0= No
Periventricular Leukomalacia (PVL)	BB05	Binary	By 30 days post-treatment	At least one AE for PVL. 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. Prolonged hospitalization of infant,

c. Persistent or significant disability/incapacity of the infant,

d. Required medical or surgical intervention to prevent any of (a) through (c) above,

e. Is considered life-threatening if no medical intervention is provided.

SAEs, treatment-related SAEs, and SAEs with an outcome of death will be summarized in the manner outlined in Section 10.2, pending there are enough events to summarize. Deaths not attributed to SAEs will also be summarized.

Note: SAEs are included in the interim safety monitoring reports. See Section 7.4.1 for additional details.

10.4 Clinical Outcomes

Select additional outcomes considered common neonatal morbidities and captured as part of the NRN Generic Database (GDB) will be reported for this study and summarized separately from the other adverse event summaries. These clinical outcomes include:

- Death in-hospital prior to 120 days chronologic age
- Necrotizing enterocolitis (NEC) defined as Modified Bell's Staging Criteria ≥ Stage 2 from birth to NRN GDB status
- Death or NEC (stage ≥ 2) from birth to NRN GDB status, composite
- Patent ductus arteriosus (PDA) from birth to NRN GDB status
- Death or PDA from birth to NRN GDB status, composite
- PDA managed with medical therapy from birth to NRN GDB status
- Death or PDA managed with medical therapy from birth to NRN GDB status, composite
- PDA managed with surgery or cardiac catherization from birth to NRN GDB status
- Death or PDA managed with surgery or cardiac catherization from birth to NRN GDB status, composite
- Severe Retinopathy of Prematurity (ROP) defined as Stage 3 ROP in either eye from birth to NRN GDB status
- Death or severe ROP from birth to NRN GDB status, composite

Most of these clinical outcomes are reported via yes/no indicators in the GDB; for those that are not, standard NRN GDB definitions will be used for their derivation.

Clinical outcomes will be listed and the number of individuals experiencing each type of event will be summarized. Tables may also include two-year historical rates of these morbidities among infants of matching gestational age born at all of the centers of the NICHD NRN.

Kaplan-Meier survival curves by treatment group will be created for the time to in-hospital death (with censoring for GDB status). Time to death will also be listed and summarized by median and range.

10.5 Growth Outcomes

Growth impairment outcomes will be collected and summarized at 36 weeks PMA, as outlined in the following table. Indicators for whether an infant's growth outcomes are below the 10^{th} percentile and associated Z-scores will be calculated based on commonly used growth charts for preterm infants.

Variable	CRF Source	Type	Collection Timepoint	Definition
Safety Outcomes				
Growth Impairment – Weight (10th pct.)	NG03	Binary	At 36 weeks PMA	Weight at 36 weeks PMA is in lowest 10th percentile. 1= Yes 0= No
Growth Impairment – Death or Weight (10th pct.)	NG03	Binary	At 36 weeks PMA	Infant died by 36 weeks PMA or weight at 36 weeks PMA is in lowest 10th percentile. 1= Yes 0= No
Growth Impairment – Weight	NG03	Continuous	At 36 weeks PMA	Z-score of weight at 36 weeks PMA
Growth Impairment – Length (10th pct.)	NG03	Binary	At 36 weeks PMA	Length at 36 weeks PMA is in lowest 10th percentile. 1= Yes 0= No
Growth Impairment – Death or Length (10th pct.)	NG03	Binary	At 36 weeks PMA	Infant died by 36 weeks PMA or length at 36 weeks PMA is in lowest 10th percentile. 1= Yes 0= No
Growth Impairment – Length	NG03	Continuous	At 36 weeks PMA	Z-score of length at 36 weeks PMA
Growth Impairment – Head Circumference (10th pct.)	NG03	Binary	At 36 weeks PMA	Head circumference at 36 weeks PMA is in lowest 10th percentile. 1= Yes 0= No
Growth Impairment – Death or Head Circumference (10th pct.)	NG03	Binary	At 36 weeks PMA	Infant died by 36 weeks PMA or head circumference at 36 weeks PMA is in lowest 10th percentile. 1= Yes 0= No
Growth Impairment – Head Circumference	NG03	Continuous	At 36 weeks PMA	Z-score of head circumference at 36 weeks PMA

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.
- Following SAS rules, the median will be reported as the average of the two middle numbers if the dataset contains even numbers.
- 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

There are no changes to the analyses planned in the protocol to date.

13 REFERENCES

Behrman RE, Butler AS, editors. Preterm Birth: Causes, Consequences, and Prevention. Washington (DC) 2007.

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.

Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatrics*. 2013;13(59):1-13.

Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics.* 1985;41:361-372.

Halvorsen T, Skadberg BT, Eide GE, Roksund OD, Carlsen KH, Bakke P. Pulmonary outcome in adolescents of extreme preterm birth: a regional cohort study. *Acta Paediatr*. 2004;93(10):1294-300. Epub 2004/10/27. PubMed PMID: 15499947.

Hascoet JM, Picaud JC, Ligi I, Blanc T, Daoud P, Zupan V, et al. Review shows that using surfactant a number of times or as a vehicle for budesonide may reduce the risk of

bronchopulmonary dysplasia. *Acta Paediatr*. 2018;107(7):1140-4. Epub 2017/12/02. doi: 10.1111/apa.14171. PubMed PMID: 29193276.

Jensen EA, Dysart K, Gantz MG, McDonald S, Bamat NA, Keszler M, et al. The diagnosis of bronchopulmonary dysplasia in very preterm infants: an evidence-based approach. *Am J Respir Crit Care Med.* 2019;200:751–759. Epub 2019/09/15. PubMed PMID: 30995069.

Jobe AH, Bancalari E. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med*. 2001;163(7):1723-9. Epub 2001/06/13. PubMed PMID: 11401896.

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

Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13-22.

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;157:940-943.

Northway WH, Jr., Moss RB, Carlisle KB, Parker BR, Popp RL, Pitlick PT, et al. Late pulmonary sequelae of bronchopulmonary dysplasia. *N Engl J Med*. 1990;323(26):1793-9. Epub 1990/12/27. doi: 10.1056/NEJM199012273232603. PubMed PMID: 2247118.

Olsen IE, Groveman SA, Lawson ML, Clark RH, Zemel BS. New intrauterine growth curves based on United States data. *Pediatrics*. 2010;125(2):e214-224.

Roberts JK, Stockmann C, Dahl MJ, Albertine KH, Egan E, Lin Z, et al. Pharmacokinetics of Budesonide Administered with Surfactant in Premature Lambs: Implications for Neonatal Clinical Trials. *Current Clinical Pharmacology*. 2016;11(1):53-61. Epub 2015/09/30. PubMed PMID: 26416605; PubMed Central PMCID: PMC5716806.

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. 1987. New York: John Wiley & Sons, Ltd.

Short EJ, Klein NK, Lewis BA, Fulton S, Eisengart S, Kercsmar C, et al. Cognitive and academic consequences of bronchopulmonary dysplasia and very low birth weight: 8-year-old outcomes. *Pediatrics*. 2003;112(5):e359. Epub 2003/11/05. PubMed PMID: 14595077.

Stevens TP, Finer NN, Carlo WA, Szilagyi PG, Walsh MC, Gantz MG, et al. Respiratory Outcomes of the Surfactant Positive Pressure and Oximetry Randomized Trial. *J Pediatr*. 2014 Apr 8. pii: S0022- 3476(14)00194-2. doi: 10.1016/j.jpeds.2014.02.054. [Epub ahead of print]. PMID24725582; PMCID4111960.

Venkataraman R, Kamaluddeen M, Hasan SU, Robertson HL, Lodha A. Intratracheal Administration of Budesonide-Surfactant in Prevention of Bronchopulmonary Dysplasia in Very Low Birth Weight Infants: A Systematic Review and Meta-Analysis. *Pediatr*

Pulmonol. 2017;52(7):968-75. Epub 2017/02/07. doi: 10.1002/ppul.23680. PubMed PMID: 28165675.

Yeh TF, Chen CM, Wu SY, Husan Z, Li TC, Hsieh WS, et al. Intratracheal Administration of Budesonide/Surfactant to Prevent Bronchopulmonary Dysplasia. *Am J Respir Crit Care Med*. 2016;193(1):86-95. Epub 2015/09/10. doi: 10.1164/rccm.201505-0861OC. PubMed PMID: 26351971.

Yeh TF, Lin HC, Chang CH, Wu TS, Su BH, Li TC, et al. Early intratracheal instillation of budesonide using surfactant as a vehicle to prevent chronic lung disease in preterm infants: a pilot study. *Pediatrics*. 2008;121(5):e1310-8. Epub 2008/04/23. doi: 10.1542/peds.2007-1973. PubMed PMID: 18426851.

Zhao W, Ciolino J, Palesch Y. Step-forward randomization in multicenter emergency treatment clinical trials. *Acad Emerg Med*. 2010;17(6):659-665.

Zhao W, Weng Y. Block urn design – a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemp Clin Trials*. 2011;32(6):953-961.

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

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
Subject Eligibility
Subject Disposition
Protocol Deviations
Study Treatment Exposure
Demographic and Baseline Characteristics
Primary Efficacy Outcome Results (overall and by GA strata)
Secondary Efficacy Results for Primary Outcome
Secondary Efficacy Outcomes Results
Number of Subjects Experiencing each AE (overall, by severity, by relationship to study drug)
Number of Subjects Experiencing each SAE (overall, fatal, related to study drug)
Mortality
Clinical Outcomes
Growth Outcomes
Figures
CONSORT Diagram
Forest plots for adverse events
Kaplan-Meier survival curves
Data Listings
None.