STATISTICAL ANALYSIS PLAN

Darbe: Darbepoetin Trial to Improve Red Cell Mass and Neuroprotection in Preterm Infants

IND: 100138

SAP VERSION: 5.0

SAP DATE: January 20, 2023

SPONSOR: NICHD

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

AUTHOR (S): Scott Grey, PhD
Dennis Wallace, PhD
Abhik Das, PhD
Marie Gantz, PhD
Sylvia Tan, MS

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	Adverse Event
BP	Blood Pressure
BPD	Bronchopulmonary Dysplasia
BSID III	Bayley Scale of Infant Development III
CBC	Complete Blood Count
CEV	Circulating erythrocyte volume
Darbe	Darbepoetin alfa
DCC	Data Coordinating Center
DSMC	Data Safety and Monitoring Committee
EDC	Electronic Data Capture
EEG	Electroencephalography
ELBW	Extremely Low Birthweight
EPO	Erythropoietin
ESA	Erythropoiesis stimulating agents
GA	Gestational Age
GEE	Generalized estimating equation
GLMM	General linear mixed models
GMF	Gross Motor Function
ICH	Intracranial Hemorrhage
IND	Investigational New Drug
IRB	Institutional Review Board
IV	Intravenous
MDI	Mental Development Index
MAR	Missing at Random
MNAR	Missing Not at Random
NDI	Neurodevelopmental Impairment
NEATO study	Neonatal Erythropoietin and Therapeutic Hypothermia Outcomes in Newborn Brain Injury (NEATO)
NEC	Necrotizing Enterocolitis
NHLBI	National Heart Lung and Blood Institute
NICHD	*Eunice Kennedy Shriver* National Institute of Child Health and Human Development
NRN	Neonatal Research Network
NS	Normal Saline
PMA	Postmenstrual age
PRBC/RBC	Packed Red Blood Cells

RCT	Randomized Controlled Trial
Retic	Reticulocyte
ROP	Retinopathy of Prematurity
RTI	Research Triangle Institute International
SAE	Serious Adverse Event
SAP	Statistical analysis plan
SAS	Statistical Analysis System
SD	Standard Deviation
SQ or SC	Subcutaneous
µg	Micrograms
VLBW	Very Low Birth Weight
WBC	White Blood Cell

1 BACKGROUND AND PROTOCOL HISTORY

Advances in neonatal care have led to significant improvements in the survival of the nearly 60,000 very low birth weight (VLBW) infants born each year in the U.S.[1] Improving neurodevelopmental outcomes for these preterm infants continues to be a major goal for neonatal care providers. A subset of these infants sustain a grade 3 or 4 intraventricular hemorrhage (IVH) resulting in an increase in the incidence of developmental delay.[2] Moreover, almost one third of preterm infants with normal head ultrasounds also develop cognitive delay.[3] Although a variety of neuroprotective treatment strategies have been evaluated, no specific treatment has been identified to reduce or prevent brain injury in these most vulnerable preterm infants.

A potential neuroprotective therapy involves administering erythropoiesis stimulating agents (ESAs) such as erythropoietin (Epo) and Darbepoetin (Darbe, a longer acting ESA). In addition to stimulating erythropoiesis, ESAs have been shown to be protective in the developing brain in animal models, making it possibly beneficial for very premature infants who are at risk for intraventricular hemorrhage, hypoxic-ischemic injury, and developmental delay.[4] Preliminary studies suggest that ESAs improve short term and preschool neurodevelopmental outcome in premature infants.[5] Prior work also indicates that Darbe appears safe within established dosing guidelines, and that the doses of Darbe administered should be adequate to achieve serum Epo concentrations ≥500 mU/mL.[6] Before treatment recommendations can be made, a well-designed RCT is needed to evaluate the effect of ESAs on various cognitive domains. Furthermore, understanding the neurologic mechanisms affected by ESAs in former premature infants is necessary to determine how best to modify therapy to achieve optimal neurodevelopmental outcome.

Applying the experience and rigor of the Neonatal Research Network infrastructure in performing randomized placebo controlled trials, our specific aims are to evaluate the effect of Darbe administered in the first 10 weeks of life to preterm infants born at 23 to 28 completed weeks gestational age (GA) on neurocognitive outcomes at 22-26 months adjusted age. Secondary outcomes include examining the impact of Darbe administration on hematocrit, red cell mass, donor exposures, number and volume of transfusions, hospital days, differences in morbidities, moderate and severe neurodevelopmental impairment (NDI), death, and cerebral palsy. If outcomes of the previous study are confirmed, the use of Darbe could become standard of care and significantly improve the lives of thousands of preterm infants.

2 PURPOSE OF THE ANALYSES

This statistical analysis plan (SAP) contains detailed information about statistical analyses to be performed to evaluate the efficacy and safety of weekly administration of Darbepoetin to preterm infants using data from a randomized, masked, placebo controlled clinical trial. The results of these analyses will be included in the clinical study report, primary manuscript and secondary manuscripts.

3 STUDY OBJECTIVES AND OUTCOMES

3.1 Study Objectives

3.1.1 Primary Study Objective

The primary objective is:

- To determine the efficacy of weekly administration of Darbe during the neonatal period to improve neurocognitive outcome at 22-26 months compared to placebo in premature infants < 28 weeks gestation.

3.1.2 Secondary Study Objectives

The secondary objectives are to determine if:

- Preterm infants administered weekly Darbe during the neonatal period will have increased red cell mass, decreased transfusions, decreased donor exposures, and decreased volume of transfused red cells compared to placebo infants.

- Preterm infants administered weekly Darbe during the neonatal period will have improved survival without NDI compared to placebo infants.

- Preterm infants with peak serum Epo concentrations ≥500mU/mL will have better neurocognitive development than those who have peak serum Epo concentrations <500mU/mL.

- Preterm infants administered weekly Darbe will have a decreased incidence of cerebral palsy compared to placebo infants.

- Preterm infants administered weekly Darbe who receive at least 1 transfusion will have improved cognitive outcome compared to placebo infants who receive at least 1 transfusion.

3.2 Outcomes

3.2.1 Primary Outcome

The primary efficacy outcome for this study is neurocognitive function at 22-26 months corrected age, measured with the Bayley Scale of Infant Development (BSID) III composite cognitive score. The BSID III is a standardized measure of development, and the raw score obtained will be converted to a standardized score based on the adjusted age for prematurity of the child. The BSID-III is well standardized, with the scores at the 22-26 months visit having a robust standard error of measure.[8] For example, an infant born at 27 weeks gestation is born 13 weeks early (approximately 3 months). This infant's follow up testing would be performed 25 months after their birthdate, which is 22 months corrected age (subtracting 3 months to correct for their prematurity). The 22-26 month window of assessment used by the

NRN realistically allows centers to find and evaluate the children within the adjusted age range the BSID III is designed to be standardized. The BSID III has been used in other neonatal trials to measure neurocognitive function, notably the neonatal ECMO study of temperature (NEST)[9] and with some NICHD NRN sites evaluating the feasibility of autologous cord blood cells for infants with hypoxic-ischemic encephalopathy;[10] in this last study, the composite score was found to be sensitive to differences in cognitive function across treatment arms. The BSID III is administered by annually certified examiners following the procedures listed in the BSID III Manual[8] and the NRN Follow-Up Study MOP.

Circumstances that inhibit the BSID III administration and scoring include the following: (1) a subject has a combination of neurologic impairment, developmental delay, blindness and/or profound hearing loss that makes the BSID III impossible to administer; (2) factors such as site staff inability to schedule a follow-up assessment with the subject, acute illness, interpreter unavailability for non-English speaking subjects, or severe behavioral problems result in the BSID III not being administered; and (3) when a subject expires prior to BSID III administration. A final circumstance is the possible inadvertent administration of the Bayley 4 exam, instead of the Bayley III; this is expected to be rare. Section 7.3 discusses how missing composite cognitive scores are handled under each of these circumstances, and section 9.3 details how a sensitivity analysis will be used to evaluate the impact of these approaches on the estimate of treatment effects.

3.2.2 Secondary Outcomes

- Hematocrit and red cell mass, also known as circulating erythrocyte volume (CEV).
- Epo concentration
- Number and volume of transfusions, number of donor exposures.
- Hospital days, defined as the number of days between date of birth and earliest of:
  - The date of discharge
  - Transfer to another hospital
  - Death
  - The infant's chronologic age if infant is still in the hospital.
- Occurrence of necrotizing enterocolitis (NEC), Bell's Stage II or worse, if treated surgically.
- Bronchopulmonary Dysplasia (BPD), using NICHD physiologic definition: requiring oxygen to maintain an oxygen saturation of ≥90% while breathing room air at 36 weeks PMA.

- Retinopathy of Prematurity (ROP), if any intervention provided.

- Intraventricular hemorrhage (IVH), classified at Grade I or higher as described by Papile et al.[7]

- Cerebral palsy, diagnosed using the NRN Follow-Up Neurological Examination, Form NF05, with three criteria: (1) abnormalities in tone, deep tendon reflexes, coordination and movement; (2) some disorder of motor function; and (3) aberrations in primitive reflexes and postural reactions that may be present.

- Neurodevelopmental impairments (NDI). For the purposes of this study, NDI will be defined as follows:

    - Severe: a BSID III cognitive score < 70, Gross Motor Functional (GMF) Level of 3-5, blindness (<20/200 vision) or profound hearing loss (inability to understand commands despite amplification);

    - Moderate: a BSID III cognitive score 70-84 and either a GMF level of 2 or a hearing deficit requiring amplification to understand commands or unilateral blindness

    - Mild: a BSID III cognitive score 70-84, or a cognitive score ≥ 85 and any of the following: presence of a GMF level 1 or hearing loss not requiring amplification.

    - Normal (no NDI) will be defined by a cognitive score ≥ 85 and absence of any neurosensory deficits.

- All cause death, defined as death from any cause following randomization. The primary, underlying cause of death will be certified by the Principal Investigator (PI) at each center who may also list co-contributing causes.

3.2.3 Safety Outcomes

Occurrence of adverse events (AEs) and serious adverse events (SAEs) observed during the study monitoring period: onset of study drug to 7 days (168 hours) after discontinuation of study drug. Study specified AEs are:

- **Major Vessel Thrombosis** defined below in various categories:

    - Moderate: Any venous or arterial thrombosis involving a major vessel.

    - Severe: Any thrombosis that is treated with a course of anticoagulation.

    - Life-threatening: Any symptomatic thrombosis involving a major vessel (e.g. symptoms such as superior vena cava syndrome)

- **Seizures (receiving treatment)** defined below in various categories:

- Moderate: Suspected, clinical, no treatment or <72 hours of treatment
- Severe: And/or confirmed (EEG) and/or treated w/ anticonvulsants ≥72 hours
- Life-threatening: Refractory seizures/ status epilepticus or decorticate posture

- **Hypertension (receiving treatment)** defined below in various categories:
  - Moderate: Systolic BP >100 repeated at least once (i.e., 2 consecutive measures) in a 24-hour period
  - Severe: Sustained increase as defined > 24 hours, treatment received
  - Life-threatening: Receiving multiple treatment types or prolonged hospitalization
- **Sepsis (culture positive)** defined below in various categories:
  - Moderate: Treated ≥ 5 days of consecutive antibiotics/virals/fungals, no cardiovascular instability
  - Severe: Treated for systemic symptoms and with ≥5 days with anti-infective agents, and additional symptomatic support needed, responded to treatment
  - Life-threatening: Same as grade III (severe) but with cardiovascular instability and slow response or no response to treatment.
- **Injection site reaction** defined below in various categories:
  - Subcutaneous (SC/SQ)
    - Moderate (SC): Hard, indurated area at site of previous injection measuring less than 1 cm in diameter
    - Severe (SC): Hard, indurated area at site of previous injection measuring 1 cm to 3 cm in diameter (any direction) or elevated
    - Life-threatening (SC): Positive blood culture thought to be related to injection site
  - Intravenous (IV)
    - Moderate IV: Moderate swelling, blanching and pain at site; good pulse and 1-2 second capillary refill time below the site, but skin cool to touch; blister may or may not be present
    - Severe IV: Severe swelling, blanching, pain, decreased or absent pulse, skin cool to touch with some evidence of necrosis (dark areas) and decreased or absent pulse below site or treated for IV infiltration per local guidelines

- Life-threatening IV: Positive blood culture thought to be related to injection site

- Other events to include:
  - Moderate: moderate illness or condition with new or significantly altered treatment-discomfort to cause some interference with usual activity.
  - Severe: severe illness or condition unresponsive to medical treatment-incapacitating with inability to perform usual activity with (usually) significant interference with normal functions.
  - Life Threatening: life threatening illness or condition; requires major surgery or respiratory support.
  - Any event leading to death

Additional details of the secondary outcomes definitions including how statistical endpoints will be developed based on these outcomes and how missing outcomes will be handled are provided in Sections 9.2, 10, and 7.3 respectively.

4 STUDY METHODS

4.1 Overall Study Design and Plan

This two-arm, parallel, randomized, masked, placebo-controlled clinical study will randomize a total of 650 preterm infants (23 0/7-28 6/7 weeks gestation) 1:1 to one of two groups: Darbepoetin 10 μg/kg/once every week (IV or SC), or placebo (equal volume normal saline for IV administration, or sham dosing) once a week, starting within 36 hours of birth, and continuing until 35 completed weeks gestation, discharge, death, or transfer to another hospital. Randomization will be stratified by center and GA. Infants in both groups will receive parenteral iron dextran or iron sucrose, 3 mg/kg once a week while they are receiving <60 mL/kg/day in enteral feedings. Monitoring for iron overload/insufficiency will occur during the study around days 14 and 42 and adjusted as needed. The primary outcome of the study, neurocognitive function at 24 months, will be obtained using the BSID III by certified examiners.

4.2 Study Population

4.2.1 Inclusion Criteria

An infant will be eligible for study participation if he/she meets the following criteria:

- Inborn and outborn preterm infants
- 23 0/7-28 6/7 weeks gestation
- ≤24 hours postnatal age

4.2.2 Exclusion Criteria

An infant will be excluded from the study if he/she meets any of the following criteria:

- Hematocrit > 60%
- Infants with known congenital or chromosomal anomalies, including congenital heart disease and known brain anomalies
- Hemorrhagic or hemolytic disease
- EEG- confirmed seizures
- Congenital thrombotic disease
- Systolic blood pressures >100 mm Hg while not on pressor support
- Receiving Epo or Darbe clinically, or planning to receive Epo or Darbe during hospitalization
- Infants in whom no aggressive therapy is planned
- Family will NOT be available for follow-up at 22-26 months

4.3 Study Arm Assignment and Randomization

After parental consent is obtained, infants will be randomized through the Data Coordinating Center (DCC) at RTI using a web based electronic data capture (EDC) system. Randomized subjects will be stratified by center and within each center by GA groups (<26 weeks vs. 26-28 and 6/7 weeks GA). Stratified randomization (1:1 Darbe or placebo) will be performed using randomly permuted blocks, with block sizes known only to the DCC. Multiple gestation infants will be randomized to the same treatment arm.

4.4 Masking and Data Lock

4.4.1 General Masking Procedures

Primary providers and bedside caregivers, parents, research personnel be responsible for data collection, and neurodevelopmental follow up personnel at enrolling sites will be masked to the treatment arm. The only persons unmasked at enrolling sites will be the pharmacists who are not otherwise involved in any other aspects of this study. Furthermore, non-site study team members including the study PI, the Darbe subcommittee for the NICHD NRN, and the sponsor (NICHD NRN) will also be masked while the study is ongoing. The DCC study statistician, responsible for reporting to the DSMC, and the lead DCC protocol coordinator, responsible for monitoring protocol compliance and SAE reporting, will be unmasked to fulfill their responsibilities. The DCC senior investigators will remain masked unless it becomes necessary to address an urgent situation at enrolling sites or to maintain the masking

of on-site personnel. Any deviations from this plan will be detailed in the clinical study report.

4.4.2 Database Lock

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 study statistician at the DCC and enrolling site pharmacists will have access to individual treatment assignment until the end of the study. Data lock and unmasking will occur at the end of the study when: (1) the last infant enrolled has had their 22-26 months corrected age neurodevelopmental evaluation or have been determined to be lost to follow-up; (2) all data quality queries from the DCC and study sites have been addressed and no further queries are anticipated; and (3) all required approvals have been obtained. After the completion of the database lock, follow-up investigators and clinical staff may be unmasked to individual treatment assignment, if requested. Additionally, parents may request and receive information about which treatment their infant received.

Any deviations from this plan will be discussed in the clinical study report. For example, the clinical study 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.

4.5 Study Flow Chart of Assessments and Evaluations

Procedures	Admission -36 hours after birth	Study Day	1-13	14	15-41	42	43-245	22-26 months corrected
		Study Week	1-2		3-7		8-35	
Screening and Informed Consent[a]	X							
CBC with reticulocyte count[b]	X			X	(X)	X		
Erythropoietin concentration[c]	X			X	(X)			
Ferritin concentration[d]				X	(X)	X		
Randomization[e]	X							
Study Drug Administration[f]	X		X		X		X	
Iron Supplementation			X		X		X	
Transfusions[g]			X		X		X	X
Adverse Event Assessments[h]	X		X		X		X	X
Follow-up Evaluation[i]								X

[a] Parents of infants with GA 23 0/7 – 28 6/7 weeks and who are not known to have any exclusion criteria may be approached for consent before delivery if allowed by the local IRB. This will enable collection of cord blood at delivery.
[b] Prior to study drug administration and at day 14 and 42.
[c] Prior to study drug administration and at day 14.
[d] Day 14 and 42.
[b-d] Clinically drawn labs [a maximum of 3 scavenged samples, marked as (X) above], in addition to the timed samples, will be collected in order to perform population pharmacokinetics.
[e] Infants who meet inclusion/exclusion criteria must be enrolled and randomized within 24 hours.
[f] The first dose of study medicine will be administered as soon as possible, at the latest by 36 hours of age.
[g] All donors, transfusions and transfusion volumes will be recorded.
[h] The following safety data will be collected through 7 days past the last study dose or to conclusion: thromboses, seizures receiving treatment, hypertension receiving treatment, culture positive sepsis, soft tissue infections at the injection site.
[i] Neurodevelopmental evaluation will be performed by certified follow up examiners at 22-26 months corrected age.

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.

Safety Population

The safety population will include all infants who were randomized and received at least 1 dose of study drug. The safety population will be used for all safety analyses and infants will be grouped per actual treatment received.

Intent-to-Treat Population

The ITT population is the primary population for formal efficacy analyses. This population includes all subjects randomized, with the exception that survivors who are lost to follow up (estimated to be less than 10% in each treatment group) will be excluded from analysis of outcomes evaluated at 22-26 months. 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.

Per-Protocol Population

The per-protocol population will be used for secondary sensitivity analyses of efficacy. This population includes all subjects who received treatment according to randomized assignment and per-protocol through 35 completed weeks gestation, 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 this study is determined entirely by comparing the Darbe and placebo groups on the primary outcome of neurodevelopmental function, measured with the BSID III composite cognitive score at 22-26 months. Comparisons of secondary outcomes between groups will be considered descriptive, and not formal tests of hypotheses. The table below presents a range of sample size estimates for each arm of the two-arm Darbe study for different underlying assumptions about the study.

Key assumptions for this study that are incorporated into each of the sample sizes calculated in the table are: (1) multiples will be randomized to the same arm; (2) 75% of the infants will survive on both arms with survival equal on the two arms and infants who do not survive will have an imputed BSID score of 54; (3) an additional 10% of infants will be lost prior to the follow-up and will be excluded from the analysis under the assumption that the data are missing at random; (4) Composite cognitive scores for survivors on the two arms will have mean values in the range of 85 and 95 with the true standard deviation will be in the range of 10 to 15 among survivors.

Sample size calculations for the Darbe study with varying effect size, standard deviation and multiple effect assumptions.

Effect Size		Standard Deviation	Multiple Effect	Sample Size Per Arm	
Survivor	Aggregate			80% Power	90% Power
7.5	5.625	17	12%	179	238
7.5	5.625	17	15%	241	322
7.5	5.625	19.5	12%	234	313
7.5	5.625	19.5	15%	241	322
7.5	5.625	22	12%	299	397
7.5	5.625	22	15%	307	408
10	7.5	17	12%	101	134
10	7.5	17	15%	104	138
10	7.5	19.5	12%	134	178
10	7.5	19.5	15%	137	183
10	7.5	22	12%	169	225
10	7.5	22	15%	174	231

Under these assumptions, the standard deviation across the mixture of survivors and non-survivors with scores imputed at 54 was found to be in the range of 17 to 20 if the underlying standard deviation among survivors was 10 and between 19 and 22 if the underlying standard deviation among survivors was 15, so three values over the range of 17 to 22 were considered. The adjustment for multiples was estimated to be in the range of 12% to 15%. Note that the final sample sizes incorporate both the multiple randomization effect and the loss to follow-up percentages. Sample sizes were computed for both 80% power and 90% power. In our previous multicenter study comparing cognitive outcomes at 18 to 22 months, we used information presented in our preliminary data section that showed a difference of 15±15 MDI points among survivors between the two groups. The differences in our current Darbe study are 8±12 points on the BSID composite cognitive score among survivors despite the relatively small sample size. We anticipate that the proposed trial will find a difference at least that large between survivors randomized to receive Darbe compared to those randomized to placebo, but the study is conservatively powered to detect a difference of 7.5 points. When non-survivors with assigned scores of 54 are included, the overall expected difference between the treatment groups becomes 5.625 points. Using a conservative estimate of differences in BSID III cognitive score of 5.625±19.5 points between Darbe recipients and controls, with 90% power and an α of 0.05, estimating a survival rate of 75% and an additional loss at follow up of 10%, and assuming multiples would be assigned to the same treatment arm, and assigning a score of 54 to non-survivors, a total of 322 (rounded to 325) infants will need to be enrolled in each arm of the study, for a total of 650 infants.

7 STATISTICAL / ANALYTICAL ISSUES

7.1 General Rules

Data will be summarized by treatment group. For summaries of study data, categorical measures will be summarized in tables listing the frequency and the percentage of subjects in each study arm; 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.

Most statistical computations will be performed and data summaries will be created using SAS 9.3 or higher. If additional statistical software is required, this will be discussed in the study report.

7.2 Adjustments for Covariates

In general, summaries and analyses will be stratified by or adjusted for gestational age strata. Specifically, table summaries will be presented for all subjects and for each gestational age strata used for randomization. All model-based analyses and test-statistics examining the treatment effect will be adjusted for study center and gestational age strata where possible. For example, the primary outcome will be tested using a linear regression implemented in a generalized estimating equations (GEE) model with an identity link function, controlling for strata defined by study center and gestational age. All other demographic and baseline characteristics for subjects will be compared between treatment groups. If analyses of these characteristics suggest that substantial differences exist for some of these characteristics between treatment arms at baseline, their use as covariates will be explored in the adjusted exploratory analyses of efficacy and safety data.

7.3 Handling of Dropouts and Missing Data

The primary analysis for publication for the overall study population as detailed in Section 9.3 will be conducted using the standardized BSID III composite cognitive score evaluated at 22-26 months. Four situations are likely to inhibit administering the BSID III to study subjects. First, when a subject has a combination of neurologic impairment, developmental delay, blindness and/or profound hearing loss, the BSID III may be impossible to administer, occurring in less than 2% of infants.[11] In this circumstance, the reason for no administration will be recorded and a score of 54 (the lowest possible BSID III composite cognitive score) will be assigned in accordance with NRN Follow-Up Study MOP procedures (section 12.1.7). This imputed value reflects the low level of neurocognitive function the study subject will likely possess as shown by the listed conditions. The second circumstance is when site staff cannot schedule a follow-up with the subject or events such as acute illness, an interpreter is not available for non-English speaking children, or severe behavioral problems preclude completion of the exam. Based on previous data, this will occur in about 8% of infants.[11] In this circumstance, the reason for no administration will again be recorded and the child will be considered missing at random (MAR), as these factors can be considered unrelated to neurocognitive development (and hence to outcome measures);

exclusion of these individuals from the analysis as MAR is unlikely to bias the treatment estimate. The third circumstance is death of the subject prior to test administration, occurring in about 16% of infants.[12] Because such a death is a post-treatment event that may be affected by treatment, treating data missing as a consequence of this event as ignorably missing could induce a biased treatment estimate on the primary outcome. While there is little evidence that Darbe or any other ESA changes mortality,[6,13] subjects who expire prior to BSID III administration will be assigned a value of 54 (the lowest possible BSID III composite score). Based on discussions with the clinical team about how they (and in their view parents) viewed death compared to severe impairment, imputation of 54 was selected as the scoring alternative that represented the least likelihood of introducing bias. While there was some disagreement among the investigators about whether death was a better, worse, or equivalent to being alive but having such neurologic morbidity that the infant achieved a lowest possible score, the general consensus was that these two conditions represented comparable levels of impairment severity. A final situation that may prevent the collection of BSID III data is that sites are commencing the administration of the BSID 4 exam in 2022, contemporaneously with Darbe follow-up. It is possible that this newer version of the Bayley exam could be mistakenly administered to Darbe subjects. We expect this situation to be rare as coordinators will be trained in the Darbe protocol and offered frequent reminders to administer the correct version of the exam. In the unlikely event that it occurs, the Bayley 4 version of the cognitive score will be used in the primary analysis and in the secondary analysis of NDI, in place of the Bayley III. Sensitivity analysis of the impact of these approaches on the estimate of treatment effects are discussed in detail in section 9.4.

Otherwise, analysis of secondary efficacy and safety data will generally include available data such that no data obtained within the study assessment windows will be discarded and no imputation for missing data will be done.

7.4 Interim Analyses and Data Monitoring

While the study is ongoing, the independent NRN Data Safety and Monitoring Committee (DSMC) will examine accumulated data to ensure protection of subjects' safety and assure that the study's scientific goals are being met. Interim analyses for this study will be primarily focused on monitoring patient safety in this extremely vulnerable population, with interim efficacy and futility analyses only attempted if at least 25% of the enrolled infants, regardless of survival, reach 2 years corrected age, presented during a scheduled safety report (as described in section 7.4.1), before study enrollment has been finished (as described in section 7.4.2). Given that the projected enrollment period for this study is less than 3 years and the relatively long follow up interval of 22-26 months corrected age, this makes it likely that only one interim efficacy and futility analysis will be possible for the DSMC to evaluate before trial enrollment is complete. However, if recruitment is unexpectedly slow, sufficient data may be available for multiple interim efficacy and futility analysis.

All interim analyses for this study 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 DSMB 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 interim safety analysis will compare the incidence of serious adverse events (SAEs) across the placebo and Darbe groups. Pre-specified formal safety looks at the interim data will occur after the first 20 patients enrolled have reached 35 completed weeks gestation, discharge, or transfer to another hospital, and subsequently, after 25%, 50% and 75% of the enrolled patients have reached the same milestone. Pocock stopping bounds will be used as stopping rules for safety, based on the four planned interim safety looks at the data. Thus, at each interim safety look, a p-value less than 0.0158 obtained from comparing the incidence of SAEs across groups, may be used by the DSMC as evidence of significant harm from the study intervention. SAEs are defined as any adverse event (defined in section 3.2.3) that 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.

See section 3.1 part iii of the Darbe Data Safety Monitoring Plan for complete definitions of SAEs. The above analysis will be conducted using robust Poisson regression implemented in a generalized estimating equations (GEE) model to adjust for both center and familial clustering, and adjusting for gestation age group to obtain the p value for comparison with the cut-off of 0.0158.

7.4.2 Efficacy

Interim efficacy looks will be performed once 25% of the enrolled infants, regardless of survival, would reach 2 years corrected age, and every 25% thereafter. These interim efficacy analyses will be presented during scheduled safety reports, conditional on the trial enrollment not having already been completed. This approach will likely permit only one interim analysis. If enrollment is slow, sufficient data may accrue for more than one interim efficacy

analyses. To appropriately control for type I error while maintaining flexibility to perform an uncertain number of interim analysis of efficacy, we will utilize a Lan-DeMets alpha spending function with an O'Brien Fleming-type stopping bound; 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 when (or if) prior analysis have taken place.[14] Utilizing SAS/STAT (version 14.2) SEQDESIGN procedure, the most likely scenario will be one interim analysis at 25% then the final analysis with corresponding p-values of 1.43×10^{-5} and 0.05. If data collection is slow and a second interim analysis is conducted, then the calculated p-values would be 1.43×10^{-5} (at 25%), 0.003 (at 50%) and 0.049 at the final analysis. The analysis will be conducted using GLMM for the primary outcome, adjusting for gestational age group and center, with a random effect for familial clustering, to obtain the p-value for comparison with the appropriate stopping boundary.

7.4.3 Futility

Interim futility analysis will be performed following the same pattern as the interim efficacy analysis, and as noted above for interim efficacy analysis, it is likely that this will permit only one interim futility analysis. However, if enrollment is slow it is possible for data to accrue for more than one interim futility analyses. In either case, the conditional power to detect a statistically significant treatment effect parameter will be calculated at each interim analysis based on the observed test statistic from accrued data, assuming the hypothesized treatment effect for the unobserved data using the two-sided conditional power calculation from Jennison and Turnbull.[15] The DSMC may recommend stopping further enrollment for futility if, at any interim analysis, the conditional power is less than 0.2.

7.5 Multicenter Studies

There are 15 NRN clinical centers taking part in this trial. While it is expected that many centers will meet or exceed recruiting the expected number of study subjects (41), smaller centers may have enrollment less than 10 subjects. Enrollment of fewer than 10 infants is expected to be rare, but if it occurs, those centers will be pooled with the nearest geographically located center in the final efficacy analysis detailed in section 9.

7.6 Multiple Comparisons and Multiplicity

As described above, formal hypothesis tests will be conducted for only one outcome for this two-group randomized trial, related to detecting a treatment effect in the primary outcome. The primary analysis for publication planned for the overall study population is described in Section 9.3. However, that outcome may be subject to multiple interim analyses, and adjustment for Type I error will be made by using a Lan-DeMets spending rule as described above in section 7.4.2.

All other analyses of outcomes are exploratory in nature; therefore, resulting p-values and confidence intervals will generally be provided for descriptive purposes only. As such, the only adjustment for multiplicity will be for the planned interim analysis.

7.7 Assessment Windows

For the primary efficacy outcome, neurocognitive function at 22-26 months corrected age is the assessment window. Otherwise, all other data will be summarized and analyzed as collected, see the listing of efficacy variables in section 9.2. Additionally, the number of assessments obtained outside of window 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 SUBJECT CHARACTERIZATION

8.1 Subject Disposition

Subject eligibility status will be summarized and listed by study arm. The number of subjects randomized; completing or discontinuing from study drug; reaching NRN status (defined as being discharged, remaining in hospital, dying or transferring to another hospital); and completing 22-26 month follow-up will be summarized by study arm. Reasons for study drug discontinuation will be listed. Additional variables to be derived, listed and/or summarized include:

- Time until NRN status by type of status event (discharge, transfer, death): Date of NRN status – Date of Birth

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 listed by center with information such as type of deviation, time of occurrence, and reason. Incidence rate of protocol deviations will also be summarized overall and for each protocol deviation category by center.

- Incidence rate of protocol deviations: number of deviations divided by the number of subject weeks at the center

8.3 Study Drug Exposure

Characteristics of study drug exposure and iron supplementation will be summarized by study arm. Characteristics include:

- Date/time of each weekly dose
- Actual dose received
- Route of administration
- Reason dose NOT received

- Doses given more than 2 days before or after a scheduled dose
- Study drug held per protocol
- Permanently stopping study drug per protocol
- Physician request to stop study drug
- Missed iron dosing >7 days
- Parental withdrawal from study with or without continued data collection

8.4 Demographic and Baseline Characteristics

Demographic and baseline characteristics for the study subjects will be summarized by study arm. Variables of interest include: Gestational age (weeks), gestational age stratum, birth weight, head circumference and length, maternal age at start of study drug, sex, race, ethnicity, use of prenatal steroids, chorioamnionitis, delivery by cesarean section, 1 and 5 minute Apgar scores, use of chest compressions or resuscitation drugs in the delivery room, and early onset sepsis (<72 hours).

- Age at start of study drug will be calculated as: Day 1 Date– Date of birth

9 EFFICACY ANALYSES

9.1 Overview of Efficacy Analyses Methods

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

9.2 Efficacy Variables

Variable	Type	When measured	Definition
Primary Outcome			
9.2.1.1 Neurocognitive function	Continuous	22-26 month follow-up	BSID III composite cognitive score, standardized based on the adjusted age for prematurity. From the Bayley III Scales Summary Score Sheet (NF09A). • If an infant dies before follow-up, a score of 54 will be assigned. • If the BSID III cannot be administered due to any combination of neurologic impairment, developmental delay, blindness and/or profound hearing loss, a score of 54 will be assigned. • If the BSID III cannot be administered due to acute illness, an interpreter is not available for non-English speaking children, or severe behavioral problems, the score will be treated as lost to follow-up (missing).

Secondary Outcomes			
9.2.1.2 Hematocrit	Continuous	Repeated over hospitalization	The ratio of the volume of red blood cells to the total volume of blood.
9.2.1.3 CEV	Continuous	Repeated over hospitalization	Circulating erythrocyte volume, calculated as hematocrit (%) x estimated blood volume (85 mL/kg) x weight (kg).
9.2.1.4 Epo concentration	Continuous	Repeated over hospitalization	Erythropoietin concentration, a glycoprotein hormone that controls red blood cell production.
9.2.1.5 Number of blood transfusions	Count	Repeated over hospitalization	Count of the number of times an infant receives a red blood cell transfusion.
9.2.1.6 Volume of blood transfusions	Continuous	End of hospitalization	The volume of all red blood cell transfusions in milliliters.
9.2.1.7 Number of donor exposures	Count	End of hospitalization	Count of the number of individuals who donate blood to infant.
9.2.1.8 Hospital days	Count	End of hospitalization	Date of birth to date infant is discharged to home, dies, or transferred to another facility
9.2.1.9 Necrotizing enterocolitis (NEC)	Binary	End of hospitalization	Medical condition where portions of the bowel undergo necrosis (tissue death). =1 Bell's Stage II or worse, if treated surgically =0 otherwise
9.2.1.10 Bronchopulmonary Dysplasia (BPD)	Binary	End of hospitalization	A chronic lung disorder common in infants with low birth weight and receive prolonged mechanical ventilation to treat respiratory distress syndrome. =1 NICHD Physiologic Definition: Requiring oxygen to maintain an oxygen saturation of ≥90% while breathing room air at 36 weeks PMA =0 otherwise
9.2.1.11 Retinopathy of Prematurity (ROP) if intervention provided	Binary	22-26 month follow-up	Abnormal growth of blood vessels in the eye, affecting prematurely born infants having received intensive neonatal care. =1 If ROP diagnosed in either eye and infant had surgery, medication or other therapies =0 otherwise
9.2.1.12 Intraventricular hemorrhage (IVH)	Binary	22-26 month follow-up	A bleeding into the brain's ventricular system. =1 from GDB NG03 Q4, cranial sonograms done within 28 days of birth with blood/echodensity in germinal matrix/subependymal area, ventricle with or without enlargement, or parenchyma =0 otherwise
9.2.1.13 Cerebral palsy	Binary	22-26 month follow-up	A group of permanent movement disorders that appear in early childhood. =1 If Neurological Examination Form NF05 Q10 is coded "Level 1" or higher =0 otherwise
9.2.1.14 Neurodevelopmental impairment (NDI)	Categorical	22-26 month follow-up	Severity of neurodevelopmental impairments, in terms of cognition, motor control, blindness or hearing loss. Severe: a BSID III cognitive score < 70, Gross Motor Functional (GMF) Level of 3-5, blindness

			(<20/200 vision) or profound hearing loss (inability to understand commands despite amplification) Moderate: a BSID III cognitive score 70-84 and either a GMF level of 2 or a hearing deficit requiring amplification to understand commands or unilateral blindness Mild: a BSID III cognitive score 70-84, or a cognitive score ≥ 85 and any of the following: presence of a GMF level 1 or hearing loss not requiring amplification Normal (no NDI) will be defined by a cognitive score ≥ 85 and absence of any neurosensory deficits
9.2.1.15 Death	Binary	Any time	=1 Death from any cause following randomization =0 if alive at 22-26 month follow-up
9.2.1.16 NDI (moderate/severe) and Death Composite	Binary	Any time	=1 moderate or severe NDI and/or Death from any cause following randomization =0 if alive and no moderate nor severe NDI at 22-26 month follow-up

9.3 Primary Analyses Methods for Scientific Publication

All analyses will be performed on an intent-to-treat basis, with the exception that survivors who are lost to follow up (estimated to be less than 10% in each treatment group) will be excluded from analysis of outcomes evaluated at 22-26 months. The primary outcome is BSID III composite cognitive score. The primary analysis (and all analyses examining outcomes by treatment) will be adjusted for the stratification variables of gestation and center. Since our primary outcome is continuous, we will use linear regression implemented in a generalized estimating equations (GEE)[16] model with an identity link function to estimate the adjusted mean difference in BSID III cognitive scaled scores between the two treatment groups; the model will include fixed effects for treatment group, gestational age group, and clinical site, and familial clustering will be accounted for through the working covariance matrix as described below. Formally, the point and interval estimates and hypothesis tests will be obtained using the following statistical model:

$$\mathbf{Y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Where $\mathbf{Y} = [Y_{ij}]^T$, with Y_{ij} being the BSID III composite score for the jth infant in the ith delivery; $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_{31}, \ldots, \beta_{3(k-1)})^T$ a matrix of parameter estimates encoded in the design matrix $\mathbf{X}$, with β_0 being the intercept parameter, β_1 the parameter estimate for the effect of Darbe compared to placebo, β_2 the parameter estimate for the effect of the high gestational age group compared to the low gestational age group, and $\beta_{31} \ldots, \beta_{3(k-1)}$ the parameter estimates for the effect of each center (pooled per the procedures stated in section 7.5) compared to the reference (kth) center, and an error term $\boldsymbol{\varepsilon} = [\varepsilon_{ij}]^T$, assumed to be distributed Normal (0, V), where V is the working covariance matrix of the BSID composite scores, encompassing both between infants mean variability in the scores as well as the covariance for these scores within infants from the same delivery.

A p-value less than alpha (with alpha dependent on the Lan DeMets spending function) for a two-sided test of $\beta_1 = 0$ will be considered as a statistically significant evidence of a treatment effect. Note that because the number of interim analyses that may be conducted is based on enrollment rates as detailed in section 7.4, the alpha at the final analysis is expected to be in the range of 0.049 to 0.05.

If the results of the primary analysis are positive, then additional analyses of the primary outcome will also be performed to examine internal consistency of the study results. Specifically, consistency of treatment effect across subsets defined by GA strata and center separately will be assessed using GEE models similar to the one used for the primary analysis. For GA strata, an interaction between GA strata and treatment will be added to the model. A p-value for the interaction >0.2 will be considered indicative of no interaction effect.

9.4 Sensitivity Analyses of Missingness in the Primary Outcome

To evaluate whether study inference is robust to both the assumptions about the missing primary outcome data and the methods for handling those missing data affect study inference, we will conduct the following sensitivity analyses. The GEE methods employed in the primary analysis assume that missing data are missing completely at random (MCAR). The robustness of the primary analysis results to potential violation of this assumption will be evaluated with four sensitivity analysis within and across each of the three anticipated situations when the BSID III cannot be administered to study subjects. Each of the sensitivity analysis below will be applied to the three anticipated missing data patterns (being too impaired, lost to follow-up, death) individually and in all subjects with a missing outcome together to assess the impact of changing the assumed missingness assumption on treatment effects.

The first sensitivity analysis will utilize multiple imputation with auxiliary variables to estimate treatment effects. This approach will employ the fully conditional specification approach[17] to generate multiple imputed data sets utilizing treatment group, gestation group, and center variables included in the primary analysis as well as eight auxiliary variables measured in first postnatal week that have been shown to predict mortality and NDI (birth weight, male gender, 5 min Apgar score, highest fraction of inspired oxygen, IVH grade, days on continuous positive airway pressure, days on conventional ventilator and days on high-frequency ventilator)[11] to increase statistical efficiency and reduce bias.[18] The number of data sets imputed will be equal to 100 times the fraction of missing outcome data (i.e. if 25% of children outcome missing data, 25 data sets will be imputed), with a minimum of 10 imputed data sets.[19] After imputation, each imputed data set will be analyzed separately using the statistical methods described in section 9.3, and parameter estimates will be combined using the Rubin variance formula to produce estimates and standard errors that incorporate missing data uncertainty.[20] This imputation approach assumes values are missing at random (MAR) meaning missing values do not depend on unobserved data, given the availability of observed outcome data and covariates. Multiple imputation will provide an evaluation of the impact of missing data uncertainty on standard errors compared to the imputation of a single value of 54 in subjects with missingness due to being too impaired or dying.

The second sensitivity analysis employs multiple imputation with a delta adjustment tipping point approach. This approach, recommended in the NRC report on missing data,[21] imposes the assumption that subjects from the Darbe treatment arm with missing data would, on average, have their unobserved BSID III scores differ by some amount δ (delta) from similar subjects with observed data in the Darbe treatment arm. A sequence of analysis assuming different δ values are performed to find "tipping points" under which the treatment effect is altered. The δ values would range between -30 and 0, in 7.5 point (half a standard deviation) increments, where delta equal zero is equal to the multiple imputation MAR analysis and each reduction in delta represents a penalty for missingness in the treatment group.[22]

The third sensitivity analysis will use a reference-based multiple imputation approach. Reference-based imputation is a pattern mixture model initially proposed by Little and Yao where the "patterns" are defined by treatment group and data from the "reference" pattern only is used in the estimation of the imputation model.[23] For this sensitivity analysis, the reference group will be children randomized to receive the placebo. Note that in this analysis, placebo missing outcomes are imputed assuming MAR, while missingness for the Darbe arm are assumed MNAR (missing not at random).[24] This effectively assumes children randomized to receive Darbe who have missing outcomes will tend to have BSID III scores similar to children in the placebo group. Thus, it is likely to provide an attenuated estimate of treatment effect if Darbe alters neurocognitive function. Implementation of the second and third sensitivity analysis is simplified by recent advancements in SAS/STAT 14.2, where the MNAR statement in PROC MI permits direct application reference-based imputation and delta adjustment on imputed BSID III scores in the treatment arm subjects.[25]

The fourth and final sensitivity analysis will employ the selection model approach to estimate treatment effects under the MNAR assumption.[26] The selection model approach assumes that it is possible to model complete outcomes (both missing and observed) jointly with a missingness indicator, and that the two models are linked by common explanatory covariates and the complete outcome data. The selection model will simultaneously use a multivariate normal model for the BSID III scores and logistic regression to model missingness probabilities, with treatment group, gestational age group, and center indicator variables and the eight auxiliary variables as covariates. Implementation of the selection model will be carried out using PROC MCMC as described by Chen, where the covariance matrix will take on an inverse-Wishart prior distribution, and the rest of the parameters will be assigned flat priors.[27]

The treatment effect estimates from each of the four sensitivity analyses will be compiled by the three anticipated missingness patterns and overall and compared to the primary analysis.

If necessary, an additional sensitivity analysis will be performed, excluding the small number of subjects that may be mistakenly administered the Bayley 4 instead of Bayley III exam from the primary analysis; results of this sensitivity analysis will be compared to the primary.

In addition to the sensitivity analyses detailed above, the secondary outcomes of mortality, neurodevelopmental impairment (NDI) and a composite of mortality and NDI will be compared across Darbe and control subjects. These analyses will inform how mortality

impacts the primary outcome across Darbe and placebo in the primary analysis and the subsequent sensitivity analyses.

9.5 Secondary Analyses Methods

Comparisons of secondary outcomes between groups will be considered descriptive, and not formal tests of hypotheses. For continuous secondary outcomes measured serially over time, including hematocrit, CEV, platelet count, and absolute reticulocyte count we will use longitudinal GLMM accounting for the lack of independence between repeated measured on the same participant to obtain estimates of the values over time in each treatment group and adjusted mean differences between the two groups.[28] Because the timing of these measures will vary by infant, days since treatment initiation will be a continuous covariate in the longitudinal models, and quadratic and cubic effects for time may be included depending on the outcome. Interactions between the time effects and treatment group will be included to assess whether the outcome measures have different trajectories of change over time in the two groups. Continuous outcomes measured at one time point, such as length of hospital stay, will be analyzed using similar GLMM models that are not longitudinal.

For categorical outcomes including NDI, death, and other morbidities, and for count outcomes such as number of transfusions and donor exposures, we will use robust Poisson regression implemented in a generalized estimating equations (GEE) model to adjust for familial clustering, and with fixed effects for center and gestational age group, to obtain adjusted relative risk estimates for the treatment effect.[16]

10 SAFETY ANALYSES

10.1 Overview of Safety Analyses Methods

All safety analyses will be performed using the safety population (i.e., as treated) unless otherwise specified. Descriptive p-values comparing the study arms will be provided on most safety table summaries and will be obtained using chi-square tests for binary outcomes specified below in Section 10.2. If the number of events allow, a 2-sided Cochran Mantel-Haenszel test controlling for strata defined by study center and gestational age will be used to obtain the p-values for binary outcomes.

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. AE will be reported and graded using the Toxicity Table for Premature Neonates: NICHD NRN (Appendix A of the Darbe Manual of Procedures). Using this table, events will be listed and 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 the MOP) 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 study arm experiencing an AE. Any events starting outside of the reportable timeframe will be included in separate listings and will be excluded from summary tables.

For the displays above, only monitored on-study AEs (listed in Section 3.2.3) 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. 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.

Table of Monitored On-Study Adverse Events

Variable	Type	Definition
10.2.1.1 Major Vessel Thrombosis	Binary	The formation of a blood clot inside a blood vessel, occurring with moderate or higher severity as indicated in Appendix A of the MOP
10.2.1.2 Seizures (receiving treatment)	Binary	An episode of signs or symptoms due to abnormal excessive or synchronous neuronal activity in the brain, occurring with moderate or higher severity as indicated in Appendix A of the MOP
10.2.1.3 Hypertension (receiving treatment)	Binary	Elevated blood pressure, occurring with moderate or higher severity as indicated in Appendix A of the MOP
10.2.1.4 Sepsis (culture positive)	Binary	An immune response triggered by an infection that can cause tissue injury, occurring with moderate or higher severity as indicated in Appendix A of the MOP
10.2.1.5 Subcutaneous (SC) Injection Site Reaction	Binary	Inflammation or damage to the tissue surrounding where the study drug is injected, occurring with moderate or higher severity as indicated in Appendix A of the MOP
10.2.1.6 Intravenous (IV) Injection Site Reaction	Binary	Inflammation or damage to the tissue surrounding the intravenous catheter where is study drug delivered occurring with moderate or higher severity as indicated in Appendix A of the MOP
10.2.1.7 Other events	Binary	Other events that are deemed either unexpected and/or at least possibly related to study or results in death, occurring with moderate or higher severity as indicated in Appendix A of the MOP

10.3 Deaths and Serious Adverse Events

An SAE is any event that is life threatening, results in death, causes or prolongs hospitalization, leads to a disability or birth defect, or requires an intervention to prevent a disability.

SAEs will be listed and SAEs, treatment-related SAEs and SAEs with an outcome of death will be summarized in the manner mentioned in Section 10.2 pending there are enough events to summarize. Separate displays listing and summarizing deaths occurring after start

of study drug through 7 days after last study drug dose including age at death and cause of death (including primary and contributing causes) will also be created.

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-value 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

Version 5.0 of the SAP incorporates a change in the planned analysis method for the primary outcome, BSID III composite cognitive score, to account for the rare possibility that the BSID 4 may be administered instead. Darbe subjects administered the BSID 4 will have that version's composite cognitive score analyzed in place of the BSID III version. Sensitivity analyses will be conducted excluding such subjects from the primary analysis.

13 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 of Drug Exposure Demographic and Baseline Characteristics 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 Primary Efficacy Results Primary Efficacy by GA Strata Sensitivity Analyses for Primary Efficacy Results Secondary Efficacy Outcomes
Figures
Hematocrit, CEV, platelet count, and absolute reticulocyte count over time by treatment arm
Data Listings
Subject Eligibility Subject Disposition Protocol Deviations Study Drug Exposure Demographic and Baseline Characteristics Reportable, On-study Adverse Events Adverse Events Not Included in Summaries Serious Adverse Events Mortality Primary and Secondary Efficacy Outcomes

14 REFERENCES

1. Meadow W, Lee G, Lin K, Lantos J. Changes in mortality for extremely low birth weight infants in the 1990s: implications for treatment decisions and resource use. Pediatrics. 2004;113:1223-1229.

2. Patra K, Wilson-Costello D, Taylor HG, Mercuri-Minich N, Hack M. Grades I-II intraventricular hemorrhage in extremely low birth weight infants: effects on neurodevelopment. J Pediatr. 2006;149:169-173.

3. Laptook AR, O'Shea TM, Shankaran S, Bhaskar B. Adverse neurodevelopmental outcomes among extremely low birth weight infants with a normal head ultrasound: prevalence and antecedents. Pediatrics. 2005;115:673-680.

4. Juul S. Erythropoietin in the central nervous system, and its use to prevent hypoxic-ischemic brain damage. *Acta Paediatr Suppl*. 2002;91:36-42.

5. Neubauer AP, Voss W, Wachtendorf M, Jungmann T. Erythropoietin improves neurodevelopmental outcome of extremely preterm infants. *Ann Neurol*. 2010 May;67(5):657-66.

6. Warwood TL, Ohls RK, Weidmeier SE, Lambert DK, Jones C, Scoffield SH, Neeraj G, Veng-Pederson P, Christensen RD. Single-dose Darbepoetin administrations to anemic preterm neonates. *J Perinatol*. 2005;25:725-30.

7. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500gm. *J Pediatr*. 1978; 92: 529-534.

8. Bayley N, 2006. *Bayley Scales of Infant and Toddler Development– Third Edition*. San Antonio, TX: Harcourt Assessment.

9. Field D Juszczak E Linsell L Azzopardi D Cowan F Marlow N Edwards D NEST Study Collaborative Group. Neonatal ECMO study of temperature (NEST): a randomized controlled trial. *Pediatrics*. 2013 Nov;132(5):e1247-56

10. Cotten CM, Murtha AP, Goldberg RN, Grotegut CA, Smith PB, Goldstein RF, Fisher KA, Gustafson KE, Waters-Pick B, Swamy GK, Rattray B, Tan S, Kurtzberg J. Feasibility of autologous cord blood cells for infants with hypoxic-ischemic encephalopathy. *J Pediatr*. 2014 May;164(5):973-979.e1.

11. Ambalavanan N, Carlo WA, Tyson JE, Langer JC, Walsh MC, Parikh NA, Das A, Van Meurs KP, Shankaran S, Stoll BJ, Higgins RD; for the Generic Database, Follow-Up Subcommittees of the Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Outcome trajectories in extremely preterm infants. *Pediatrics*. 2012 Jul;130(1):e115-25.

12. Stoll BJ1, Hansen NI, Bell EF, Shankaran S, Laptook AR, Walsh MC, Hale EC, Newman NS, Schibler K, Carlo WA, Kennedy KA, Poindexter BB, Finer NN, Ehrenkranz RA, Duara S, Sánchez PJ, O'Shea TM, Goldberg RN, Van Meurs KP, Faix RG, Phelps DL, Frantz ID 3rd, Watterberg KL, Saha S, Das A, Higgins RD; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Neonatal outcomes of extremely preterm infants from the NICHD Neonatal Research Network. *Pediatrics*. 2010 Sep;126(3):443-56.

13. Fauchère JC, Koller BM, Tschopp A, Dame C, Ruegger C, Bucher HU; Swiss Erythropoietin Neuroprotection Trial Group. Safety of early high-dose recombinant erythropoietin for neuroprotection in very preterm infants. *J Pediatr*. 2015 Jul;167(1):52-7.e1-3.

14. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13:1341-1352.

15. Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*, New York: Chapman & Hall, 2000. 210-213.

16. Hardin J, Hilbe J, 2003. *Generalized Estimating Equations*. London: Chapman and Hall/CRC.

17. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007; 16(3):219-242.

18. Meng XL. Multiple imputation with uncongenial sources of input. *Stat Sci*. 1994; 9:538–574.

19. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011; 30:377–399.

20. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). Wiley: Hoboken, NJ, 2002.

21. National Research Council. Panel on Handling Missing Data in Clinical Trials, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press: Washington, DC, 2010.

22. Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials. *J Biopharm Stat*. 2009; 19:1085–1098.

23. Little RJA, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*. 1996; 52:471–483.

24. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharm Stat.* 2013; 12: 337–347.

25. Yuan Y. Sensitivity analysis in multiple imputation for missing data. *Proceedings of the SAS Global Forum 2014 Conference*: http://support.sas.com/resources/papers/proceedings14/SAS270-2014.pdf.

26. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *Applied Stat*. 1994;43:49–73.

27. Chen F. Missing no more: using the PROC MCMC procedure to model missing data. *Proceedings of the SAS Global Forum 2013 Conference*: http://support.sas.com/resources/papers/proceedings13/436-2013.pdf.

28. Fitzmaurice GM, Laird NM, Ware JH, 2004. *Applied Longitudinal Analysis*. John Wiley & Sons, Inc.

29. McLean RA, Sanders WL, Stroup WW. A unified approach to mixed linear models. *Am Stat*. 1991;45(1): 54–64.