A Randomized Trial of Targeted Temperature Management with Whole Body Hypothermia for Moderate and Severe Neonatal Encephalopathy in Premature Infants 33-35 Weeks Gestational Age - A Bayesian Study

NCT01793129

Statistical Analysis Plan

January 3, 2023

STATISTICAL ANALYSIS PLAN

A RANDOMIZED TRIAL OF TARGETED TEMPERATURE MANAGEMENT WITH WHOLE BODY HYPOTHERMIA FOR MODERATE AND SEVERE NEONATAL ENCEPHALOPATHY IN PREMATURE INFANTS 33-35 WEEKS GESTATIONAL AGE

A BAYESIAN STUDY

Short title: Preemie Hypothermia

SAP VERSION: Version 1.1

SAP DATE: 1/3/2023

PI/SPONSOR: Roger G. Faix, NICHD

PREPARED BY: RTI International

AUTHOR (S): Barry Eggleston

Contents

1 ADMINISTRATIVE INFORMATION

1.1 Trial Name

1.1.1 Trial Registration

ClinicalTrials.gov Identifier: NCT01793129

1.2 Protocol Version

Version 3.0

1.3 SAP Revision History

Ver.	Justification for change	Date
0.1	Initial Version	2021-05-3
0.2	2nd draft	2021-10-05
0.3	3rd draft	2021-12-03
0.4	4th draft	2022-06-13
0.6	6th draft	2022-09-13
0.7	7th draft	2022-10-18
1.0	Approved	2022-11-02
1.1	Per study subcommittee deliberation over email on March 3rd and 4th, 2023, changed the model used for analysis of ordinal outcomes from a proportional odds model to a set of logistic regression models with post processing to facilitate calculations of relative risks and/or risk differences. This change made the analysis more consistent with the analysis of previous hypothermia trials.	2023-01-02

1.4 List of Abbreviations

Abbreviation:	Definition:
AE	Adverse event
BPD	Bronchopulmonary dysplasia
DCC	Data Coordinating Center
DSMB	Data and safety monitoring board
FAS	Full analysis set
FU	Preemie Hypothermia Follow-up database
GA	Gestational age
HIE	Hypoxic-ischemic encephalopathy
ICH	Intracranial hemorrhage
hr	Hours
SC	Steering committee
IRB	Institutional Review Board
ITT	Intent-to-treat
IVH	Intraventricular hemorrhage
mITT	Modified Intent-to-treat
N/A	Not Applicable
NE	Neonatal encephalopathy
NICHD	National Institute of Child Health and Human Development
NIH	National Institute of Health
PP	Per Protocol
RS	Research site
SAE	Severe adverse event
SAP	Statistical analysis plan
Tes	Temperature per esophageal probe
wks	Weeks

2 INTRODUCTION

The Preemie Hypothermia Trial was a randomized multicenter trial conducted to determine whether whole body hypothermia for 72 hrs in preterm infants 33-35 wks gestational age (GA) and ≥ 1500 grams birth weight who present at <6 hrs postnatal age with moderate to severe neonatal encephalopathy (NE) formerly called hypoxic-ischemic encephalopathy (HIE) is safe and will reduce death or moderate/severe disability at 18-22 months corrected age.

2.1 Background and Rationale

Most clinical studies of NE and potential interventions have targeted infants ≥36 wks GA. Although many interventions have been suggested and assessed for prevention or palliation of NE in high-income countries, the only one currently supported by rigorous clinical evidence to improve outcome in human newborns has been hypothermia (targeting core temperature of 33.5 with a range of 33.0 to 34.0°C. - with or without head cooling) implemented at ≤6 hrs of postnatal age and maintained for 72 hrs in infants ≥36 wks GA with moderate or severe NE. Translation of these findings to infants who are more premature has been suggested as a subject for further scrutiny by some and actually implemented by others.

Reports of induced hypothermia include infants of lower GA, but the incidence of such infants and their outcomes are unclear. The protocol lists several reported studies and verbally reported cases where hypothermia treatment was used at GAs less than 36 wks. These cases are occurring even though the literature lacks the evidence to support hypothermia treatment at GA < 36 wks. Concern about increased frequency or severity of adverse events or other problems in infants 33-35 wks GA subjected to 'therapeutic' hypothermia is real, but there are minimal human data to determine whether benefits will outweigh risks. Given the ongoing use of induced hypothermia for indications that are not evidence-based (including use at <36 wks GA) and the unknown risk-benefit ratios, it was deemed important to complete this trial as designed to systematically generate prospective data and analyze the data in a rigorous manner.

2.2 Study Objectives

2.2.1 Study Hypotheses

The risk of death or moderate or severe disability at 18-22 months corrected age will be decreased in infants 33 0/7 - 35 6/7 wks GA and ≥ 1500 grams birth weight with moderate or severe NE due to presumed perinatal hypoxia-ischemia at < 6 hours age who undergo targeted temperature management with whole body hypothermia to 33.5°C (as monitored by esophageal temperature [Tes]) for 72 hours than in those with core temperature (also per Tes) targeted at 37.0ºC (controls).

Primary objectives

1) To assess differences in death or moderate/severe disability at 18-22 months in enrolled infants randomized to whole body hypothermia for 72 hours, compared to those randomized to normothermia for 72 hours.

2) To determine short-term safety (i.e., within 108 hours of baseline [time of insertion of esophageal probe]) of targeted temperature management with whole body hypothermia for 72 hours in infants 33 0/7 -35 6/7 weeks GA with moderate or severe NE at ≤ 6 hours of age. Safety criteria will include serious adverse effects (SAEs): intracranial hemorrhage, cardiac arrhythmia; persistent acidosis;

thrombosis; bleeding; skin changes; necrotizing enterocolitis (NEC); spontaneous intestinal perforation (SIP); esophageal perforation attributable to the placement of the esophageal tube, ulceration or bleeding from the esophageal probe; hypo- and hyperglycemia; receipt of ECMO; thrombocytopenia, and death.

Secondary objectives

To assess differences between study groups with respect to a select set of secondary efficacy and safety outcomes which are listed in Sections 6.2 and 6.3.

3 STUDY METHODS

This statistical analysis plan (SAP) contains detailed information about statistical analysis to be performed to assess the above mentioned primary and secondary objectives once the final Preemie Hypothermia analysis datasets are constructed.

3.1 Trial design

The trial was a staged prospective, randomized multicenter trial with qualifying infants 33 0/7 -35 6/7 weeks GA to be randomized to whole body hypothermia (33.5 °C esophageal temperature [Tes]) for 72 hours or normothermia (with steps incorporated to avoid hyperthermia [>37.3°C Tes]). The treatment assignment was unmasked. The "staged" component of the trial refers to how the original trial design included an assessment of enrollment after 2 years of enrollment. If enrollment was sufficient during the first 2 years, then the study would continue to enroll until completion unless the DSMB deemed it necessary to stop the study at any DSMB assessment.

3.2 Study Intervention and Process

Infants were randomized within 6 hours of age to either whole body cooling or normothermia for 72 hours. The targeted temperature management for the whole body cooling group included a targeted temperature of Tes 33.5°C, (acceptable Tes 33.0 – 34.0°C): the targeted temperature management for normothermia included a targeted temperature range of Tes 36.5 – 37.3°C, with steps to prevent Tes >37.3°C). In both groups the treatment period was 72 hours. Hypothermia is attained and maintained with the Cincinnati Sub-Zero Hyper/Hypothermia Device. Surveillance for safety and adverse events within 108 hours of baseline (insertion of esophageal probe) is conducted. This includes cranial ultrasound no later than 24 hours after baseline read by local center readers. All survivors undergo brain MRI after the intervention at 10-17 days of age, with strong encouragement to obtain the MRI after the intervention or prior to NICU discharge/transfer to non-NRN center (if <10-17 days of postnatal age). MRI studies obtained at other times post-hypothermia/control intervention are included as well, and all enrolled babies undergo standard interdisciplinary neurodevelopmental assessment at 18-22 months corrected age (although the Covid pandemic occasionally required deferral of this visit, in which case the visit was scheduled as early as possible once follow-up visits were possible).

3.3 Randomization

After informed consent was obtained, qualifying infants were randomized by 6 hours of age to either hypothermia or the non-cooled control group. Randomization occurred 24 hours a day, 7 days a week through the Data Coordinating Center at RTI International, Research Triangle Park, NC. Infants were stratified by center and degree of NE (moderate versus severe). In the unlikely event of a multiple gestation qualifying for inclusion, all who qualified were randomized independently.

3.4 Sample Size

During the design phase, the potential number of eligible subjects was uncertain due to limited and variable data such that traditional sample size determination was not likely to be helpful. The original NRN hypothermia trial in infants ≥ 36 weeks GA found a frequency of 62% death or moderate/severe disability in control infants; there were no data regarding the incidence of death or disability after NE at the targeted gestational age of this study during the design phase of this trial. Given increased fragility associated with this degree of prematurity, it seemed likely that the incidence of death or moderate/severe disability in the control arm was likely to be at least 70%.

To achieve 80% power to detect a decrease in death or moderate/severe disability from 70 to 50% a sample size of about 110 subjects per arm is needed. A decrease from 70 to 55% would also be clinically significant but would require about 190 subjects per arm. These sample size requirements were deemed unfeasible for the population involved in this research, so the Preemie Hypothermia study was designed to randomize 168 babies and analyze the data using a Bayesian approach to estimate the effect of hypothermia relative to normothermia with as much precision as possible and allow for the calculation of a probability of treatment benefit for hypothermia. The 168 was derived from projected ability to enroll among participating sites and experience from the previous Late Hypothermia trial. This trial was not powered in the traditional sense, but simulations under specified scenarios were run and the results showed that the trial design had more than a 75% chance of observing a final posterior probability of at least 0.80 for RR < 1 (indicating treatment benefit) given the data, when the true RR is close to 0.70.

3.5 Analysis Structure

The main goal of this analysis is to evaluate whether preemie infants 33-35 weeks GA with moderate to severe NE at ≤ 6 hrs age targeted to a core temperature of 33.5°C (as monitored by esophageal temperature) for 72 hours in the hypothermia arm versus core temperature targeted to 37.0°C reduces the risk of death or moderate or severe disability at 18-22 months corrected age. Secondary goals of analysis will be completed to more fully understand the effect of hypothermia on components of the primary outcome, as well as safety outcomes.

3.6 Statistical Interim Analysis and Stopping Guidance

Interim statistical analyses and stopping guidance are given in sections 4.3 – 4.6 of protocol, summarized in Table 1 of protocol.

3.7 Timing of Interim and Final Analysis

Final analyses were planned to occur after 168 infants were consented, enrolled, randomized, and followed to 18-22 months corrected age. The study analysis database is constructed using those who died before the 18-22 month follow-up and all survivors not lost to follow-up who underwent protocol-specified neurodevelopmental assessment.

4 STATISTICAL PRINCIPLES

The Preemie Hypothermia trial has a Bayesian design with Bayesian interim analyses of safety and efficacy. Interim analyses of safety and efficacy were completed as described in the protocol.

4.1 Modeling Processes, Prior Specification, and Reporting Procedures

- For binary outcomes, the protocol specified that a Bayesian log-binomial regression model with random center effect would be fit to estimate the relative risk of hypothermia relative to normothermia. However, since the protocol was written, an increasing emphasis has been placed on including inference on risk differences in clinical trial reports, so a logistic regression model with random center effect will be fit followed by post-processing of individual predicted probabilities of outcome to estimate posteriors of relative risk and risk differences.[1]
    - Weakly informative priors will be used during analyses to better understand the effect of hypothermia on outcomes considered in this trial:
        - On the relative risk and risk difference scales of primary outcome and components, the priors for the treatment effect will include a skeptical prior, a neutral prior, and an enthusiastic prior. For all other outcomes, only the neutral prior will be used.
    - For the intercept and the encephalopathy terms, broad neutral priors will be used.
    - For the random center effect, a Normal prior centered at 0 will be used where the standard deviation parameter of this prior will have a half-normal hyperprior with mean equal to 0 and standard deviation equal to 1.
    - Further details about these priors are given in section 6.4.2 and 6.4.3
    - Posterior distributions, posterior means, posterior probabilities of benefit, and 95% credible intervals will be reported for the effect of hypothermia on the outcomes considered.
- For count outcomes such as days or durations, a negative binomial model with random center effect will be fit to estimate relative risk.
    - Weakly informative priors will be used during analyses to better understand the effect of hypothermia on outcomes considered in this trial:
        - On the relative risk scale, the priors for the treatment effect will include a neutral prior.
    - For the intercept and the encephalopathy terms, broad neutral priors will be used.
    - For the random center effect, a Normal prior centered at 0 will be used where the standard deviation parameter of this prior will have a half-normal hyperprior with mean equal to 0 and standard deviation equal to 1.
    - Further details about these priors are given in section 6.4.2 and 6.4.3

[1] Andrew Gelman and Iain Pardoe. Average Predictive Comparisons for Models with Nonlinearity, interactions, and variance components. *Sociological Methodology*. May 18, 2007.

- Posterior distributions, posterior means, posterior probabilities of benefit, and 95% credible intervals will be reported for the effect of hypothermia on the outcomes considered.

4.2 Protocol violations/Deviations

Protocol violations (PV) and minor protocol deviations (MPD) were collected using a specified set of descriptions. Possible reasons for PV and MPD were:

- Study intervention never started
- Wrong treatment intervention applied
- Infant ineligible
- No consent
- Neuro exam not done
- Wrong stratification variables used at randomization
- Too early removal of esophageal probe
- Infant randomized after 6 hours
- Intervention initiated after 6 hours and 30 minutes
- Intervention discontinued early
- Equipment malfunction
- Cranial ultrasound completed outside of protocol-specified age window
- Tepid sponge bath not provided
- Infant not rewarmed per protocol
- Other violation

PVs and MPDs will be descriptively summarized in terms of infants per arm with any PV(MPD) and in terms of number of PV(MPD) per arm.

4.3 Analysis Datasets

This analysis will use two analysis populations, a full analysis set (FAS) and per-protocol (PP).[2] The FAS analysis population will include all randomized infants with ascertained primary outcomes. This FAS analysis population is recognized as being as close as possible to an intent-to-treat population. All analyses based on the FAS will analyze infants as randomized. The PP analysis population will include all infants who received their randomized treatment (72 hours of hypothermia or normothermia) and have a determined primary outcome. All analyses based on the PP will analyze infants as treated.

[2] ICH E9: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9-statistical-principles-clinical-trials

5 TRIAL POPULATION

5.1 Screening population

All infants born at 33 0/7 to 35 6/7 weeks gestation and ≥ 1500 grams are screened for study entry if they are admitted to the NICU with an admitting diagnosis of acute perinatal asphyxia, neonatal depression, encephalopathy and/or fetal acidemia at less than 6 hours of age. All these infants at < 6 hours of age are evaluated by clinical and biochemical criteria and further assessed by neurological examination. Both inborn and outborn infants are evaluated. From this population, babies are screened for inclusion in the trial using the following eligibility criteria.

Eligibility

Inclusion criteria

Infants 33 0/7 to 35 6/7 weeks GA (best obstetrical estimate) and ≥1500 grams birth weight (to minimize potential difficulties placing esophageal probe) who meet clinical, biochemical and neurologic criteria for moderate to severe NE in this GA range at ≤ 6 hours age:

1) Biochemical:

Cord gas or if unavailable, blood gas within first hour of life with pH ≤7.00 or base deficit (BD) ≥16 mEq/L, OR

2) Clinical in the absence of blood gas results or Clinical with more modest fetal acidemia (cord or 1st hour of life blood gas with a pH of 7.01-7.15 or BD > 10.0 to 15.9mEq/L):

a) Acute perinatal event (e.g., abruptio placenta, cord prolapse, uterine rupture, maternal cardiac or respiratory arrest, severe FHR abnormality such as variable or late decelerations,) AND

b) Requirement for positive pressure ventilation for apnea OR poor respiratory effort since birth for at least 10 minutes OR 10-minute Apgar score ≤ 5

AND

3) Neurologic:

Clinical Seizures OR modified Sarnat score with moderate or severe abnormalities in at least 3 of the 6 categories; at least one must be altered level of consciousness (lethargy or stupor/coma) as determined by a certified examiner.

All infants who meet criteria for potential inclusion should undergo standard neurologic exam as for infants ≥ 36 wks GA being considered for hypothermia, with findings recorded.

Criteria for exclusion in hypothermia trial

1) Receipt of paralytic agent or sufficient sedative or analgesic agent that is considered by the examiner to confound the qualifying neurologic exam

2) Etiology of NE not likely to be hypoxic-ischemic in origin

3) Major congenital anomaly that may confound outcome

4) Considered to be moribund and will not be receiving full intensive care

5) Equipment and/or appropriate staff not available

6) Infant was cooled prior to arrival at the NRN center and had a core temperature <34.0º C for more than 1 hour at the time of screening (Cooling on transport is to be discouraged, in general)

7) Unable to randomize by 6 hours of age

8) Infant to receive or receiving ECMO

9) All blood gases (cord and neonatal at <1 hour of age) have a pH >7.15 and a base deficit <10 mEq/L

10) Consent not obtained from parent(s)

11) Concurrence not provided by responsible attending neonatologist

5.2 Withdrawal / Follow-up

- Infants can be lost-to-follow-up between Network Status (120 days after birth, or time of discharge/transfer) and the time of the scheduled 18-22-month follow-up visit (although the Covid pandemic did require some subjects to undergo follow-up at later dates). The Preemie Hypothermia PH13 form is completed at the time of network status. Infants cannot be lost prior to network status.
- Infants who withdraw from the study will not have any data collected beyond the timepoint of withdrawal. Data collected prior to withdrawal will be available for summarization and analysis.
- Numbers of infants who withdraw or are lost to follow-up will be presented in the CONSORT diagram.

5.3 Baseline Patient Characteristics

Baseline characteristics will be compared across treatment groups. Table 1 in the MockTablesAppendix.docx document lists out maternal and neonatal characteristics that will be summarized in the primary published paper. In addition to the variables indicated in Table 1 of the MockTablesAppendix.docx document, other baseline characteristics may be investigated internally by the Preemie Hypothermia Subcommittee. For programming purposes, the following list contains an example list of infant and maternal characteristics that will or may be investigated as well as dataset and variable names needed to locate the characteristics in the clinical database:

- Infant baseline characteristics (*Variable Type, Form, Variable name in clinical database*)
    - Birth year (*Multiple level discrete, PH05: A.1.a, PH5BRTDT*)
    - GA (*Multiple level discrete – at level of weeks, PH05: A.7.a, PH5GESTW*)
    - Sex (*Binary, PH05: A.8, PH5SEX*)
    - Birthweight (grams) (*Continuous, PH05: A.4, PH5BRTWT*)
    - Length (cm) (*Continuous, PH05: A.5, PH5BRTLN*)
    - Head circumference (cm) (*Continuous, PH05: A.6, PH5HCIRC*)
    - Seizures (*Binary, PH02: C.1, PH2SIEZ*)
    - Level of Encephalopathy (*Binary, PH02: G.4, PH2LECEP*)
    - Inborn/outborn status (*Binary, PH05: A.2, PH5OUTBN*)
    - 1-minute Apgar scores (*Continuous, PH05: A.3.a, PH5APG1*)

- 5-minute Apgar scores (*Continuous, PH05: A.3.b, PH5APG5*)
- Received resuscitation at delivery (*Binary, PH05: B.1-B.5, PH5OXYG / PH5BAGM / PH5CHCOM / PH5INTUB / PH5DRUGS. any equal to 'Y'*)
    - Types of resuscitation (*Multiple level discrete or individual binary, use PH05: B.1-B.5, PH5OXYG / PH5BAGM / PH5CHCOM / PH5INTUB / PH5DRUGS.*)
- Received resuscitation at 10 minutes (*Binary, PH05: C.1-C.5, PH5OXY10 / PH5BAG10 / PH5CHE10 / PH5INT10 / PH5DRG10. Any equal to 'Y'*)
    - Types of resuscitation (*Multiple level discrete or individual binary, use PH05: C.1-C.5, PH5OXY10 / PH5BAG10 / PH5CHE10 / PH5INT10 / PH5DRG10.*)
- Received cord blood gas (*Binary, PH05: E.1, PH5CBGO, will be used in the construction of a derived variable containing cord blood gas information if non-missing and <1 hour of age information if cord blood gas information is missing and <1 hour of age information is non-missing. Stated another way: select cord gas if both cord gas and 1 hr gas are available and select 1 hr gas if cord not available. These variables are an intermediate step. See cord or postnatal variables below for variables that will actually get used in reports.*)
    - Source (*Multiple level discrete, PH05: E.1.a, PH5CBS*)
    - pH (*Continuous, PH05: E.1.b, PH5CBPH*)
    - $PC0_2$ (mmHg) (*Continuous, PH05: E.1.c, PH5CBPCO*)
    - PO_2 (mmHg) (*Continuous, PH05: E.1.d, PH5CBP02*)
    - HCO_3 (mEq) (*Continuous, PH05: E.1.e, PH5CBHCO*)
    - Base deficit (*Continuous, PH05: E.1.f, PH5CBBD*)
- Postnatal blood gas obtained at <1 hour of age (*Binary, PH05: E.2, PH5PNBGO, will be used in the construction of a derived variable containing cord blood gas information if non-missing and <1 hour of age information if cord blood gas information is missing and <1 hour of age information is non-missing. Stated another way: select cord gas if both cord gas and 1 hr gas are available and select 1 hr gas if cord not available. These variables are an intermediate step. See cord or postnatal variables below for variables that will actually get used in reports*)
    - Source (*Multiple level discrete, PH05: E.2.a, PH5PGS*)
    - pH (*Continuous, PH05: E.2.d, PH5PGPH*)
    - $PC0_2$ (mmHg) (*Continuous, PH05: E.2.e, PH5PGPCO*)
    - PO_2 (mmHg) (*Continuous, PH05: E.2.f, PH5PGPO2*)
    - HCO_3 (mEq) (*Continuous, PH05: E.2.g, PH5PGHCO*)
    - Base deficit (*Continuous, PH05: E.2.h, PH5PGBD*)
- Cord or postnatal (<1 hour of age) blood gas (*Combination of appropriate blood gas variables from the above two blood gas groups. Use Cord variable if non-missing and use <1 hour variable if cord variable is missing and <1 hour variable is non-missing. Stated another way: select cord gas if both cord gas and 1 hr gas are available and select 1 hr gas if cord not available. These variables will actually be the variables used in reports covering cord or postnatal blood gas.*)
    - pH (*Continuous*)
    - $PC0_2$ (mmHg) (*Continuous*)
    - PO_2 (mmHg) (*Continuous*)
    - HCO_3 (mEq) (*Continuous*)

        - Base deficit (*Continuous*)

- Maternal baseline variables
    - Mother's age (*Continuous, PH04: A.1, PH4MAGE*)
    - Mother's race (*Multiple level discrete, PH04: A.2, A.2.a.1-A.2.a.3, PH4MRACE, PH4MRAC1, PH4MRAC2, PH4MRAC3*)
    - Ethnicity (*Multiple level discrete, PH04, A.3, PH4METHN*)
    - Marital status (*Binary, PH04: A.4, PH4MARST*)
    - Education status (*Multiple level discrete, PH04: A.5, PH4EDSTA*)
    - Parity (*Multiple level discrete, PH04: A.6.b, PH4PARIT*)
    - Multiple Birth (*Binary, PH04: A.7, PH4MLTIB*)
    - Prenatal care (*Binary, PH04: A.8, PH4PRECR*)
    - Medical insurance (*Multiple level discrete, PH04: A.9, PH4INSUR*)
    - Complications during pregnancy (*Binary, PH04: B.1-B.4, PH4HYPRE / PH4ANTHM / PH4THYRM / PH4DIABT. any equal to 'Y'*)
        - Hypertension/pre-eclampsia/eclampsia (*Binary, PH04: B.1, PH4HYPRE*)
        - Antepartum hemorrhage (*Binary, PH04: B.2, PH4ANTHM*)
        - Thyroid malfunction (*Binary, PH04: B.3, PH4THYRM*)
        - Diabetes (*Binary, PH04: B.4, PH4DIABT*)
    - Complications at delivery (*Binary, PH04: D.1 through D.14, PH4DEFHT, PH4CORDM, PH4CORDO, PH4UTRPT, PH4SHDYS, PH4PLACP, PH4PLACO, PH4MATHM, PH4MAFEH, PH4MATTR, PH4MATCR, PH4MATSZ, PH4PYREX, PH4CHORI, PH4HISCH, PH4ANTIB. any equal to 'Y'*)
        - Decelerations and/or loss of fetal heart tones (*Binary, PH04: D.1, PH4DEFHT*)
        - Cord mishap (*Binary, PH04: D.2, PH4CORDM*)
        - Cord mishap, type (*Multiple level discrete, PH04: D.2.a-D.2.d, PH4CORD1 / PH4CORD2 / PH4CORD3 / PH4CORD4 / PH4CORDO*)
        - Uterine rupture? (*Binary, PH04: D.3, PH4UTRPT*)
        - Shoulder dystocia? (*Binary, PH04: D.4, PH4SHDYS*)
        - Placental problems? (*Binary, PH04: D.5, PH4PLACP*)
        - Placental problems, type (*Multiple level discrete, PH04: D.5.a-D.5.dPH4PLAC1 / PH4PLAC2 / PH4PLAC3 / PH4PLAC4*)
        - Maternal hemorrhage? (*PH04: D.6, PH4MATHM*)
        - Maternal-fetal hemorrhage? (*PH04: D.7, PH4MAFEH*)
        - Maternal trauma? (*PH04: D.8, PH4MATTR*)
        - Maternal cardio-respiratory arrest? (*PH04: D.9, PH4MATCR*)
        - Maternal seizures? (*PH04: D.10, PH4MATSZ*)
        - Pyrexia >= 37.6? (*PH04: D.11, PH4PYREX*)
        - Documented chorioamnionitis? (*PH04: D.12, PH4CHORI*)
        - Antibiotic for suspected/confirmed infection? (*PH04: D.14, PH4ANTIB*)
    - Rupture of membranes (*Binary, PH04: E.1, PH4RUPT*)
    - Duration of rupture of membranes (Continuous, birth date as date/time variable {PH05: A.1.a + A.1.b, PH5BRTDT and PH5BRTTM} minus date/time of rupture of membranes: PH04: E.1.a + E.1.b, PH4RUPDT and PH4RUPTM)

- Duration of rupture of membranes, ≤ 18 hr (Binary, based on duration of rupture of membranes
- Mode of delivery (*Multiple level discrete, PH04: F.1, PH4FNMOD*)

Note: The amount of missing data will be indicated in the Tables.

6 STATISTICAL ANALYSIS

6.1 Primary Efficacy Outcome

The primary outcome will be death or moderate to severe disability at 18-22 months corrected age (or later as required by Covid pandemic) (*Binary*).[3] Death will be identified by finding an indication of death on the PH13 network status form, PH15 adverse event form, DEATH_FU and DEATH_NEW variables from the follow-up (FU) analysis dataset, JFSTATUS = 3 from the FU analysis dataset.[4]

The presence or absence of disability will be determined by the standard NRN interdisciplinary follow-up exam. Where possible, the primary outcome values will be taken from the FU analysis dataset. Binary variables for severe, moderate, mild, and no disability will be created using the following criteria. A single binary moderate or severe disability variable will be created from this individual disability severity variable to identify infants with moderate or severe disability (= 1, if severe disability or moderate disability is indicated; = 0, if disability can be determined and not equal to severe or moderate.)

Severe disability will be defined by the presence of any of the following:

- a Bayley III cognitive score < 70 (*FU: B3_COG70*),
- Gross Motor Functional Classification Status (GMFCS) Level of 3-5 (*FU:GMF_LEV >= 3*),
  - Note: The GMF_LEV variable will need to be programmed in accordance with GMF_LEV as programmed in the regular RTI FU (follow-up) analysis dataset.
- blindness (visual acuity < 20/200 in the both eyes as determined by an ophthalmologist) (*FU: VISIONR and FU:VISIONL equal to 4 or 5, B.1.e*) or
- profound hearing loss (inability to understand commands despite placement of a cochlear implant or amplification) (*FU:HEARIMP = 1 (B.2.b) and FU:HEARAID = 1, 2, or 3 (B.2.b.1) and FU:HEARCOC = 1, 2, or 3 (B.2.b.2)*).

Moderate disability will be defined by:

[3] The primary outcome calculation contains variables that are only available in the Preemie Hypothermia Follow-up Database. When programming the primary outcome and other outcomes derived from the Preemie Hypothermia Follow-up database, you may come across outcomes that are missing. If outcomes are missing, determine what variable in the derivation is causing the missingness and then review data, regarding the infant with a missing outcome, from the PF11, PF12, and PF09A forms. The PF11 form will tell you if particular data is available, and the PF12 form will have information about lost-to-follow-up and additional information that may help understand the condition of the infant. Review the data in these forms to better understand why the data is missing and document what you find in a text file.

[4] In this SAP, FU refers to the Preemie Hypothermia FU database.

- a Bayley III cognitive score between 70-84 (*FU.B3_COG70 = 0 and FU.B3_COG85 = 1)*
  and either
- a GMFCS level of 2 (*FU.GMF_LEV = 2*), or
- an active seizure disorder requiring ongoing anticonvulsant therapy and not attributable to another disease process (*PH02: C.1, PH2SIEZ = 'Y' or PF04 A.4 = 'Y'*), or
- a hearing deficit requiring amplification or cochlear implant to understand commands (*FU.HEARCOC = 1,2, or 3 or FU.HEARAID = 1, 2, or 3*).

Mild disability will be defined by:

- a cognitive score 70-84 (*FU.B3_COG70 = 0 and FU.B3_COG85 = 1)*,
  or
- a cognitive score ≥85 (*FU.B3_COG70 = 0 and FU.B3_COG85 = 0*)

  and

- any of the following:
  - GMFCS level I (*FU.GMF_LEV = 1*),
  - seizure disorder (*PH02: C.1, PH2SIEZ = 'Y'*) or
  - hearing loss not requiring amplification (*FU:HEARIMP = 1 (B.2.b) and FU:HEARAID = 0 and FU:HEARCOC = 0 (B.2.b.2))*.

Normal will be defined by:

- a cognitive score ≥85 (*B3_COG70 = 0 and B3_COG85 = 0)* and
- the absence of any neurosensory deficits (*PH02 Q C.1. = 'N', FU.GMF_LEV = 0, FU.BLIND_06 = 0, FU.DEAF_06=0, FU.HEARAID=0, FU.HEARCOC = 0, FU.HEARIMP=0*)).

In addition to the above 4 levels of disability, a separate disability variable will be created to identify infants who have "Profound disability". Profound disability will be indicated by the assignment of lowest possible Bayley score due to inability to test due to severity of impairment. Note: any infant with the profound disability variable equal to yes will also have the primary outcome variable equal to yes, since profound disability is a subclass of severe disability with respect to the primary outcome.

Although the Bayley III cognitive score is thought to systematically give higher scores at 18-22 months than the Mental Developmental Index of the Bayley II used in the original NRN hypothermia trial for infants ≥36 weeks GA, this is likely to affect both study groups in this study similarly. Bayley III was used for the Optimizing Cooling Strategies RCT. Please note that the follow-up for this trial is at 18-22 months corrected age in keeping with all previously performed RCTs of hypothermia for HIE in newborn infants 36 weeks gestation or beyond, although the Covid pandemic required some visits to be deferred to later times.

6.2 Safety Outcomes

The safety outcomes that will be reported in the published primary paper are given in Table 2 of the MockTablesAppendix.docx document. A list of safety outcomes that may be considered in analysis are:

- Intra-cranial hemorrhage (ICH), up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- cardiac arrhythmia, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- persistent acidosis, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- thrombosis, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- major bleeding, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- Perforations, ulceration or bleeding from the esophageal probe, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- alteration of skin integrity, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)[5]
    - Erythema
    - Scherema
    - Cyanosis
    - Subcutaneous fat necrosis (during or after the intervention)
- hypoglycemia, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- hyperglycemia, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- receipt of ECMO, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- PPHN, up to 108 hours after baseline, make sure this variable is only used in analysis/reports concerned with the "during intervention" time period. (Binary, *PH15: A.1, PH15TYPE*)
- thrombocytopenia, up to 108 hours after baseline (*Binary, PH15: A.1, PH15TYPE*)
- necrotizing enterocolitis, up to 108 hours after baseline (NEC as SAE) (*Binary, PH15: A.1, PH15TYPE*)
- SIP (Binary, PH15 G.5 = 'Y' and date G.5.a = birthdate)
- Other SAE
- Clinical Seizures during intervention (*Binary, PH13: D.1.b = 'Y'*)
- Electrographic seizures (Binary, PH13: D.2.a = 'Y', and date/time D.2.b within end of intervention period)
- New ventilation requirement during 72 hr (Binary, Use PH08 respiratory support information for 24 Hr, 48 Hr, or 72Hr time points)

6.3 Hospital Course Outcomes

Hospital Course outcomes that may be considered in analysis are:

- Proven NEC (Binary, PH13, G.4, if PH13PNEC = 2 or 3.)
- Spontaneous Intestinal perforation without proven NEC (SIP) (*Binary, PH13: G.5, PH13GAST,* if PH13GAST = 'Y')

[5] Subcutaneous fat necrosis may occur after 108 hours.

- seizures after beginning of assigned treatment {Clinical Seizures} (*Binary, PH13: D.1.b, D.1.c, D.1.d, D.1.e, taken from PH13SZAB, or PH13SZDM, or PH13SZRW, or PH13SZNT equal to 'Y'*)
- culture-positive bloodstream or infection of other normally sterile site at >3 days (*Binary, PH13: E.1, E.2, PH13SEPT or PH13MENG needs to be 'Y'*)
- persistent pulmonary hypertension (PPHN) after start of intervention period (*Binary, PH13: B.2, PH13PPHN*)
- metabolic abnormalities after start of intervention period (*Binary, PH13: I.1, I.2, I.3, I.4, I.5, I.6, I.7, PH13HYGL = Y' or PH13HYRG = 'Y' or PH13HYCA = 'Y' or PH13HYMA = 'Y' or PH13HYKA = 'Y' or PH13HYRK = 'Y' or PH13HYPH = 'Y'*)
- treatment with vasopressors and/or steroids after start of intervention period (*Binary, PH13: C.7, PH13INOV*)
- oliguria after start of intervention period (*Binary, PH13: F.1, PH13OLIG*)
- anuria after start of intervention period (*Binary, PH13: F.2, PH13ANUR*)
- liver dysfunction after start of intervention period (AST>200 IU and/or ALT>100 IU) (*Binary, PH13: G.7, PH13EAST*)
- liver dysfunction after start of intervention period (elevated direct bilirubin) (*Binary, PH13: G.8, PH13EDBR*)
- clinically diagnosed coagulopathy after start of intervention period (*Binary, PH13: H.1, PH13DIC*)
- days receiving oxygen/mechanical ventilation (*Continuous, PH13: B.6, B.7, B.8, based on PH13DVEN and PH13DOXY, and PH13DCPA*)
- Ventilation (Binary, if days receiving oxygen/mechanical ventilation > 0 then = yes)
- duration of NICU stay among survivors (*Continuous, PH13: R.1, PH13DSDT (date of discharge) minus PH05: A.1.a, PH5BRTDT (birthdate)*)
- Bowel perforation (multi-level, ='NEC' if PH13 G4 = 2, ='NEC with Perforation' if PH13 G4 = 3, = 'SIP if PH13 G5 = 'Y')
- neurological injury by cranial ultrasound within 24 hours of enrollment, identified by local reader (*Binary, PH12, see sub-section Analysis of Ultrasound, CT scan, MRI and encephalopathy outcomes under section 6.5.5.*)
- Abnormal neurological examination at discharge
- Death prior to discharge
- Electrographic seizures (Binary, PH13: D.2.c = 'Y' or D.2.d='Y', and date/time D.2.b consistent with date occurring after intervention period)
- Anti-seizure medications at discharge (Binary, PH13: D.4 = 'Y')
- DNR order (Binary, PH13: P.2 = 'Y')
- DNR order and support withdrawn (Binary, PH13: P.2 = 'Y' and PH13 P.1.b.1, 2, 3, 4, or 5 = 'Y'

Note: The above listed safety outcomes will be summarized under the heading "during study intervention", while the hospital course outcomes will be summarized under the heading "during hospital course", which will be defined as the period between birth and death or discharge. See tables shell documents for more details.

6.4 Secondary Outcome

Table 3 of the MockTablesAppendix.docx document contains the secondary outcomes that will be summarized in the primary published paper. Secondary outcomes that may be considered during initial analysis will include but not necessarily limited to:

- death (*Binary, sources used to identify are listed in Section 6.1*)
- cause of death (withdrawal of support and reasons for such will be tracked) (*Multiple level discrete, PH13 PH13DCOD.*)
- disability in survivors (*Multi-level discrete,* None/mild/moderate/Severe *disability, derivations given in Section 6.1*)
- death or profound disability (assignment of lowest score on Bayley III because untestable due to degree of impairment) (*Binary, sources used to identify death are listed in Section 6.1, will need to use NF9a.nf9acsur variables to identify infants who had Bayley III cognitive score set to lowest score due to severe delay in development.*)
- survival with no disability (*Binary, no indication of death and disability assessment equal 'Normal' as calculated in Section 6.1*)
- death or severe disability (Binary, *indication of death and indication of severe disability assessment as calculated in Section 6.1)*
- moderate or severe disability among survivors (*Binary, use information regarding death and section 6.1 for calculation*)
- severe disability among survivors (*Binary, use information regarding death and section 6.1 for calculation*)
- moderate disability among survivors (*Binary, use information regarding death and section 6.1 for calculation*)
- profound disability among survivors (*Binary, use information regarding death and section 6.1 for calculation*)
- mild disability among survivors (*Binary, use information regarding death and section 6.1 for calculation*)
- no disability among survivors (*Binary, use information regarding death and section 6.1 for calculation*)
- Cerebral palsy (Any, Quadriplegic, Diplegic, Hemiplegic, Dystonic, Athetotic, Ataxic, Other, Disabling (moderate or severe), *(Binary, using information from PF05 C.9.a, C.9.b, C.9.c.)*
- a Bayley III language score < 70 (*Binary, among survivors, see section 6.1*)
- a Bayley III language score 70 - 84 (*Binary, among survivors, see section 6.1*)
- a Bayley III language score >= 85 (*Binary, among survivors, see section 6.1*)
- a Bayley III motor score < 70 (*Binary, among survivors, see section 6.1*)
- a Bayley III motor score 70 - 84 (*Binary, among survivors, see section 6.1*)
- a Bayley III motor score >= 85 (*Binary, among survivors, see section 6.1*)
- Head circumference < 10^{th} percentile, among survivors (*Binary, based on PF05 follow-up form question A.3*)
- Weight < 10^{th} percentile, among survivors (*Binary, based on PF05 follow-up form question A.1*)
- Length/height < 10^{th} percentile, among survivors (*Binary, based on PH05 follow-up form questions*)
- each component of severe and moderate disability (among survivors):

    - a Bayley III cognitive score < 70 (*Binary, see section 6.1*)
    - a Bayley III cognitive score 70 - 84 (*Binary, see section 6.1*)
    - a Bayley III cognitive score >= 85 (*Binary, see section 6.1*)
    - Gross Motor Functional (GMFCS) Level of 3-5 (*Binary, see section 6.1*)
    - blindness (visual acuity < 20/200 in the best eye as determined by an ophthalmologist) (*Binary, see see section 6.1*)
    - profound hearing loss (*Binary, see see section 6.1*)
    - a hearing deficit requiring amplification or cochlear implant to understand commands (*Binary, see see section 6.1*)
    - Seizures since discharge (*Binary, PF04 A.4 = 'Y'*)
- Abnormal neurological examination at discharge
- differences in MRI findings after cessation of cooling/control (*See subsection "Analysis of Ultrasound, CT scan, MRI and encephalopathy outcomes" under section 6.5.5 for more details*).

6.5 Analysis Methods

6.5.1 Statistical Methods

Patient accounting will be presented in a consort diagram. This diagram will include:

- Number of infants screened
    - Reasons for non-eligible
- Number of infants eligible
    - Reasons for non-consent
- Number of infants consented
- Number of infants randomized
- Number of infants randomized per group
    - Number withdrawn
    - Number lost-to-follow-up
- Number of infants who died
    - Before hospital discharge
    - Between hospital discharge and scheduled follow-up visit
- Number of infants with primary outcome per group

Baseline characteristics and outcomes will be descriptively summarized in the following manner:

Binary (yes/no 1/0 type) measures will be summarized by level with frequency out of the total non-missing values and percentage of total non-missing values. Percentages will be reported to one decimal place. For example: nn / NN (XX.X%)

Multiple level discrete measures will be tabulated with one row per level and with a row indicating how many observations had missing data. Percentages will be calculated as percent of all observations associated with a given treatment.

Continuous measures will be summarized by number of non-missing values, means and standard deviation [n: xx.x (xx.x)], or by median, minimum and maximum [n: xx.x, (xx.x, xx.x)] , or by median,

1st quartile, 3rd quartile [n: xx.x, (xx.x, xx.x)]. Standard deviations and continuous measures of mean, median, minimum, maximum, or quartiles will be reported to one decimal place.

Count measures that have more than 10 levels will be treated as continuous measures. Count measures that have 10 or fewer levels will be treated as multiple level discrete measures.

Statistical modeling and inference will be made using logistic models for binary outcomes, proportional odds models for ordinal outcomes, and negative binomial models for day and duration outcomes. All modeling will use Bayesian techniques to estimate the posterior distribution for all model parameters. Bayesian models will be fit using MCMC software such as PROC MCMC, JAGS, or Stan. The preferred software will be PROC MCMC, but if JAGS or Stan is used it will be accessed through R programming language-based packages such as rjags, rstan, rstanarm, or brms.

Model parameter estimates will be generated by fitting three MCMC chains. Each chain will involve at least 10,000 burn-in samples and at least 200,000 additional samples, which will be thinned by at least 5. More samples and more thinning will be done if necessary, to produce good sampling of the posterior. The final MCMC chains will have at least 120,000 useable samples to estimate the posterior distributions necessary for making inference about the treatment effects, (relative risk or outcome difference).

For binary outcomes, the outcome for the jth infant from the ith center will have the following form:

$$y_{ij} = \begin{cases} 0 & \text{survival without moderate / severe disability at 18-22 months corrected age} \\ 1 & \text{death or moderate / severe disability at 18-22 months corrected age} \end{cases}$$

The main covariates will have the form:

$$trt_{ij} = \begin{cases} 0 & \text{normothermia} \\ 1 & \text{whole body hypothermia} \end{cases}$$

$$enceph_{ij} = \begin{cases} 0 & \text{mild encephalopathy} \\ 1 & \text{moderate or severe encephalopathy} \end{cases}$$

The logistic model used for analysis of binary outcomes on the relative risk and risk difference scale will have the following form:

$$y_{ij} \sim Bernoulli[\pi_{ij}], \text{ where}$$
$$\text{logit}[\pi_{ij}] = \alpha_i + \theta_1 \cdot enceph_{ij} + \theta_2 \cdot trt_{ij}$$

β_1, β_2 will have priors as specified in Section 6.5.3

α_i will be the random intercepts due to center. Prior specified in Section 6.5.3.

The prior for the random center effects will be Normal centered at 0, and the hyperprior for the standard deviation of the random center effect will be specified in Section 6.5.3. Any baseline covariates that are added to the model (in sensitivity analyses to adjust for important unbalanced baseline covariates) will either be coded using 0/1 indicator variables, sets of indicator variables, or as centered continuous covariates. Priors for these other to-be-determined covariates are specified in Section 6.5.3. Post processing will be used to get posterior approximations of relative risk and risk differences due to treatment. The logistic model will be fit to the data, then for each infant, the model parameter chains will be used to estimate the within subject posterior relative risk and risk difference which will equal hypothermia divided by normothermia and hypothermia minus normothermia respectively. A within subject estimate of relative risk will be calculated by setting

the treatment variable value to hypothermia, estimating the probability of outcome for all infants, repeating the same process after setting the treatment variable to normothermia then dividing the two probabilities of outcome for all infants, hypothermia over normothermia. A within subject estimate of risk difference will be calculated by setting the treatment variable value to hypothermia, estimating the probability of outcome for all infants, repeating the same process after setting the treatment variable to normothermia then subtracting the two probabilities of outcome for all infants, hypothermia minus normothermia. Since the parameter chains will contain at least 120,000 samples, each infant will contribute at least 120,000 samples for the approximation of the within subject relative risk posterior and the within subject risk difference posterior. Using all samples of relative risk and risk differences from all infants to approximate the posteriors, the mean estimate of relative risk and risk difference will be calculated as well as, 95% credible intervals of mean relative risk and risk difference. Finally, posterior probabilities of benefit due to hypothermia from both relative risk and risk differences will be used as summary statistics. Differences in the relative risk and risk difference posteriors will be reflected upon to better understand the estimated treatment effect.

For count outcomes like days or duration, y_{ij}, the main covariates will have the form:

$$trt_{ij} = \begin{cases} 0 & \text{normothermia} \\ 1 & \text{whole body hypothermia} \end{cases}$$

$$enceph_{ij} = \begin{cases} 0 & \text{mild encephalopathy} \\ 1 & \text{moderate or severe encephalopathy} \end{cases}$$

The model used for analysis of day and duration outcomes will have the following form:

$$y_{ij} \sim NegBinomial\left[\vartheta_{ij}\right], \text{where}$$
$$ln\left(\vartheta_{ij}\right) = \alpha_i + \theta_1 \cdot enceph_{ij} + \theta_2 \cdot trt_{ij}$$

β_1, β_2 will have priors as specified in Section 6.5.4

α_i will be the random intercepts due to center. Prior specified in Section 6.5.4.

The prior for the random center effects will be Gaussian centered at 0, and the hyperprior for the standard deviation of the random center effect will be specified in Section 6.5.4. Any baseline covariates that are added to the model (in sensitivity analyses) will either be coded using 0/1 indicator variables, sets of indicator variables, or as centered continuous covariates. Priors for these other to-be-determined covariates are specified in Section 6.5.4.

The analysis of ordinal outcomes will use sets of logistic regression models as used for binary outcomes. One of the ordinal outcome levels will be designated as the reference level, and then logistic regression models will be fit to subsets of the data where one group of included data will be records that equal to reference level and the other group of included data will be records that equal one of the other outcome levels. For example, if an ordinal outcome has 3 levels, with level 3 designated as the reference level, then two logistic regression models will be fit:

- One model will use data from level 1 and level 3
- The second model will use data from level 2 and level 3

The prior for the random center effects will be Normal centered at 0, and the hyperprior for the standard deviation of the random center effect will be specified in Section 6.5.3. Any baseline covariates that are added to the model (in sensitivity analyses) will either be coded using 0/1 indicator variables, sets of indicator variables, or as centered continuous covariates. Priors for these other to-be-determined covariates are specified in Section 6.5.3.

6.5.2 Covariates

All regression models will adjust for:

- level of encephalopathy at randomization and
- center (as a random effect).

The planned number of enrolling centers is 18, and center will be included in the linear component of the model as an additive random effect.

In addition to adjusting for level of encephalopathy and center as a random effect, sensitivity analyses will be completed by selecting baseline covariates that differ across the two treatments. These covariates will be selected from those covariates listed in section 5.3.

6.5.3 Prior Specification (Binary/Ordinal Primary and Secondary Outcomes)

For the intercept and the encephalopathy terms, Normal(mean = 0, sd=1) priors will be used.

For any other covariates added to the model for sensitivity analyses or any exploratory model, a Normal(mean=0, sd=1) prior will be used.

For the random center effect, we will assume a Normal distribution centered at 0, and the standard deviation parameter will be assumed to follow a half-normal hyperprior with mean equal to 0 and standard deviation equal to 1.

For the term representing the log odds ratio of hypothermia treatment, skeptical, neutral, and enthusiastic priors will be used to estimate posterior distributions for odds ratio when analyzing the primary outcome and primary outcome components. For all other outcomes, only a neutral prior will be used. Because the logistic model will estimate the log odds ratio as the treatment parameter, the priors for the treatment effect will be normal priors on the log odds ratio scale. These normal priors for the log odds ratio are defined below so that the resulting priors in the odds ratio scale have specified properties.

On the odds ratio scale, the skeptical prior for odds ratio will be centered on 1.1 and have a 95% credible interval with an approximate range of 0.28 to 4.40. This skeptical prior will be defined by a normal prior on the log odds ratio scale centered at log(1.1) = 0.0953 with standard deviation (SD) of 0.7072, Normal(mean = 0.0953, sd=0.7072).

On the odds ratio scale, the neutral prior for the odds ratio will be centered on 1.0 and have a 95% credible interval with an approximate range of 0.25 to 4.00. This neutral prior will be defined by a

normal prior on the log odds ratio scale centered at 0 with standard deviation (SD) of 0.7072, Normal(mean = 0, sd=0.7072). Similar priors will be used for the regressions involved in proportional odds modeling of secondary ordinal outcomes.

On the odds ratio scale, the enthusiastic prior for the odds ratio will be centered on 0.75 and have a 95% credible interval with an approximate range of 0.19 to 3.00. This enthusiastic prior will be defined by a normal prior on the log odds ratio scale centered at log(0.75) =-0.2877 with standard deviation (SD) of 0.7072, Normal(mean = -0.2877, sd=0.7072).

6.5.4 Prior Specification (Secondary Count Outcomes)

For the intercept and the encephalopathy terms, Normal(mean = 0, sd=1) priors will be used.

For any other covariates added to the model for sensitivity analyses or any exploratory model, a Normal(mean=0, sd=1) prior will be used.

For the random center effect, we will assume a normal distribution centered at 0, and the standard deviation parameter will be assumed to follow a half-normal hyperprior with mean equal to 0 and standard deviation equal to 1.

For the term representing the log mean ratio of hypothermia a neutral prior will be used to estimate posterior distributions for the mean ratio. Because the negative binomial model will estimate the log rate ratio as the treatment parameter, the priors for the treatment effect will be normal priors on the log scale. These normal priors for the log rate ratio are defined below so that the resulting priors in the mean ratio scale have specified properties.

On the mean ratio scale, the neutral prior for the mean rate will be centered at 1.0 and have a 95% credible interval with an approximate range of 0.33 to 3.0. This neutral prior will be defined by a normal prior on the log mean ratio scale centered at 0 with standard deviation (SD) of 0.5605, Normal(mean = 0, sd=0.5605).

6.5.5 Final Analysis

The final analysis will occur after 168 infants have completed the trial, as specified earlier. The final analysis will not occur until all enrolled infants have a primary outcome determination, i.e. have the primary endpoint value assigned, classified as withdrawn without primary outcome assigned, or deemed lost to follow-up. Part of determining primary outcome of enrolled infants will involve masked adjudication of outcomes for infants who have an indeterminable outcome based on partially collected follow-up information. Once all enrolled infants have reached final status and primary endpoints determined, the clinical database will be exported, and study analysis datasets created. At that point, the analysis datasets will be considered “locked” and any further adjustments to the clinical datasets will not be included in the analysis datasets.

The final analysis will involve posterior distribution estimation for the effect of hypothermia relative to normal cooling. For binary outcomes, the effect will be the adjusted relative risk and adjusted risk differences calculated from a binomial regression model with logit link and post processing of

predicted outcome probabilities. For count outcomes, the effect will be the adjusted relative risk from the two groups obtained using negative binomial regression. For ordinal outcomes, the effect will be the adjusted odds ratio calculated from the cumulative logits.

For treatment relative risk estimation, posterior probabilities of >0%, >5%, >10% and >20% decrease in binary outcomes will be calculated. If θ is the relative risk defined as risk in cooling group over risk in control group, then the above specified posterior probabilities are, $Pr(\theta < 1 \mid X)$, $Pr(\theta < 0.95 \mid X)$, $Pr(\theta < 0.9 \mid X)$, $Pr(\theta < 0.8 \mid X)$. In addition to posterior probability calculations, graphical displays of the posterior distribution and appropriate 95% credible intervals for the over-all relative risk will be created. Finally, forest plots of the center level relative risks and 95% credible intervals will be created.

In addition to adjusting for level of encephalopathy and center in the logistic model, the final analysis of the primary outcome will also perform sensitivity analyses concerning possible treatment effect confounders present at randomization. These sensitivity analyses will adjust for baseline covariates that differ across the two treatments in addition to adjusting for level of encephalopathy and center. The additional covariates will be determined by comparing the primary outcome with respect to baseline characteristics listed in Section 5.3. More details about these sensitivity analyses are given in section 6.5.7. Secondary analyses may also adjust for covariates, but in such a case the analysis will only be completed with covariate adjustment. No secondary analyses will be completed twice, in the sense of one analysis without covariates and another analysis with covariates. These analyses will use the FAS population.

The lack of preexisting data suggests that the neutral prior distribution may be most appropriate for presentation of final results; however, generation of final results of the primary outcome analysis using the skeptical (RR of 1.10) and enthusiastic priors (RR of 0.75) will also be produced for the primary outcome to give a broad overview of the implications of the trial's efficacy results. Skeptical, neutral, and enthusiastic priors will also be utilized in the analysis of primary outcome components. All non-primary or non-primary component outcomes will be analyzed using only the neutral prior.

For treatment risk difference estimation, posterior probabilities of >0%, >1%, >2%, >3%, and >5% decrease in binary outcomes will be calculated (corresponding to NNTs of 100, 50, 33, and 20). If ψ is the risk difference defined as risk in normothermia group minus risk in cooling group, then the above specified posterior probabilities are, $Pr(\psi < 0 \mid X)$, $Pr(\psi < 0.01 \mid X)$, $Pr(\psi < 0.02 \mid X)$, $Pr(\psi < 0.03 \mid X)$, and $Pr(\psi < 0.05 \mid X)$. In addition to posterior probability calculations, graphical displays of the posterior distribution and appropriate 95% credible intervals for the over-all risk difference will be created. Finally, forest plots of the center level risk differences and 95% credible intervals will be created.

In addition to adjusting for level of encephalopathy and center in the logistic model, the final analysis of the primary outcome will also perform sensitivity analyses concerning possible treatment effect confounders present at randomization. These sensitivity analyses will adjust for potential treatment effect confounders in addition to adjusting for level of encephalopathy and center. The additional covariates will be determined by comparing the primary outcome with respect to baseline characteristics listed in Section 5.3. More details about these sensitivity analyses are given in section 6.5.7. Secondary analyses may also adjust for covariates, but in such a case the analysis will only be completed with covariate adjustment. No secondary analyses will be completed twice,

in the sense of one analysis without covariates and another analysis with covariates. These analyses will use the FAS population.

The lack of preexisting data suggests that the neutral prior distribution may be most appropriate for presentation of final results of risk differences; however, generation of final results of the primary outcome analysis using the skeptical (OR of 1.10) and enthusiastic priors (OR of 0.75) will also be produced for the primary outcome to give a broad overview of the implications of the trial's efficacy results. Note, these priors are expressed on the OR scale, because the underlying model is a logistic regression model and posteriors of risk difference will be generated via post-processing. Skeptical, neutral, and enthusiastic priors will also be utilized in the analysis of primary outcome components. All non-primary or non-primary component outcomes will be analyzed using only the neutral prior.

For treatment rate ratios of count outcomes estimation, the posterior of the treatment rate ratio, ϕ, will be graphically illustrated and summarized by calculating the median treatment rate ratio, the 95% credible interval of treatment rate ratio, and $Pr(\phi < 1 \mid X)$ using neutral priors. These analyses will use the FAS population.

Primary Analysis of Primary Outcome

With respect to the primary outcome, the data will be analyzed using the FAS population.

In summary the following analyses will be completed on the primary outcome:

Population	FAS					
Treatment effect	RR			RD		
Prior*	E	N	S	E	N	S

* E = Enthusiastic, N = Neutral, S = Skeptical

Each of the six analyses of the primary outcome will generate the following results:

- Posterior plot of treatment effect
- Posterior median and posterior standard deviation
- 95% credible interval
- For relative risk, posterior probabilities of >0%, >5%, >10%, and >20% decrease in the primary outcome will be calculated
- For risk difference, posterior probabilities of >0%, >1%, >2%, >3%, and >5% decrease in primary outcome will be calculated
- forest plot of the center level effect and 95% credible intervals.

If the FAS analysis of the primary outcome results in any clinically notable findings, then the primary outcome may also be analyzed in an exploratory fashion using the PP population.

Analysis of Secondary Outcomes

With respect to secondary outcomes, analyses will follow the same process as the primary outcome analysis defined above. See MockTablesAppendix.docx for list of selected variables.

Analysis of Safety (AE) and Hospital Outcomes

Safety and hospital outcomes listed in Sections 6.2 and 6.3 will be analyzed using the FAS population and if necessary, the PP population using relative risk models, proportional odds models, or Negative Binomial models, depending on the type of outcome.

With respect to binary safety outcomes, the estimated posterior distribution of relative risk, θ, based on the data and the neutral prior will be used to determine posterior probabilities of harm from the intervention, $Pr(\theta > 1 \mid X)$, as well as 95% credible intervals.

For ordinal safety outcomes, the estimated posterior distribution odds ratio for the cumulative logits, ϕ, will be estimated based on the data and the neutral prior. In addition, posterior probabilities of harm, $Pr(\phi > 1 \mid X)$, and 95% credible intervals of ϕ will be calculated.

For count type safety outcome, the estimated posterior distribution of the treatment mean rate, ϕ, will be estimated based on the data and the neutral prior. In addition, posterior probabilities of harm, $Pr(\phi > 1 \mid X)$, and 95% credible intervals of ϕ will be calculated.

Analysis of Ultrasound, CT scan, MRI and encephalopathy outcomes

Ultrasound, CT scan, MRI, and encephalopathy outcomes will be analyzed using the FAS population. Head ultrasounds, CT scans, and brain MRIs will be summarized by treatment and time (prior to intervention, during intervention, after intervention, any time). Summaries will be created for number of babies with any abnormal findings during time period data recorded, and for total events across all babies during any time period.

Normal/Abnormal findings in Ultrasound, CT scan, and MRI after intervention will be summarized by treatment. The summarized measure will be any normal/abnormal findings after intervention. In addition to normal/abnormal findings, the following abnormal findings will also be summarized by treatment with respect to any findings after intervention:

1) basal ganglia, thalamic abnormality, posterior limb of the internal capsule abnormality (based on MRI, conditions 3, 24 on PH12 form. Posterior limb of the internal capsule abnormality will be captured with condition 22 and other.)
2) white matter abnormality (based on MRI, condition 23 on PH12 form.)
3) IVH (based on MRI, conditions 5 and 6)
4) post-hemorrhagic ventricular dilatation or non-hemorrhagic ventricular dilatation (based on MRI, conditions 5, 6, and 17 on PH12 form.)
5) IVH (based on Ultrasound, conditions 5 and 6)
6) Any bleeding (based on MRI, conditions 4-10, 13, 14, 25, or 26 on PH12 form.)

Additional analyses of these outcomes will be covered under a separate SAP.

6.5.6 Assumption Checks

Convergence of MCMC chains will involve visual comparisons of the trace plots from the 3 chains generated for each model. Quantitative checks such as the Geweke and Gelman/Rubin tests will be completed to ensure convergence to the posterior.

6.5.7 Sensitivity Analysis

In addition to adjusting for level of encephalopathy and center as a random effect, sensitivity analyses may be completed, at the discretion of the Preemie Hypothermia subcommittee, by selecting baseline covariates that differ across the two treatments. These covariates will be selected from those covariates listed in section 5.3. This analysis will only be completed for the primary outcome (death or moderate to severe disability at 18-22 months corrected age) and Death. These sensitivity analyses will be based on the FAS population.

In addition to the FAS population-based analysis, exploratory PP analyses may be done using the PP population on the primary or secondary outcomes at the discretion of the Preemie Hypothermia subcommittee, if the FAS analysis of the primary or secondary outcome results in any clinically notable findings.

6.5.8 Subgroup Analysis

The primary outcome will be reanalyzed for subgroup exploratory analyses. These analyses will be completed using the FAS population. These subgroup analyses will utilize interactions within the risk difference model defined in section 6.5.1 to study treatment heterogeneity and to estimate effects within subgroups when the treatment heterogeneity is clinically important. In each case, the subgroup variable will be added to the model, if not already in the model, along with the interaction between the subgroup variable and the treatment variable. These subgroup analyses will use modified versions of the models that included potentially important covariates and described in sections 6.5.1 and 6.5.2. Only the neutral priors will be considered in these analyses, and the priors for the interaction terms will be the same priors applied to the treatment groups as described in sections 6.5.3 and 6.5.4. In these models with interaction terms, effect coding will be used for treatment and level of encephalopathy instead of reference cell coding.

Reporting of subgroup analyses will include summary statistics from the Posterior distribution of the treatment effect within each level of the subgroup as well as generation of forest plots and posterior plots when deemed useful by the Preemie Hypothermia subcommittee. In addition to estimation of Posterior treatment effects within each level of the subgroup, summary posterior statistics of differences /ratios of treatment effects from all possible combinations of subgroup level pairs will be estimated (possible calculations here will be: differences in differences or ratio of ratios, since they will represent or be functions of effects that make up the interaction term(s) and other combinations not used to create interaction terms in the model given model parameterization). Summary posterior statistics will include medians, 95% credible intervals, and posterior probabilities of benefit.

The subgroups that will be considered are:

- Sex
- Center, if model estimation will allow it
- Race, provided we have sufficient diversity in the study population. At least two levels making up at least 20% each, and at least three levels represented.
- GA (3 levels: 33, 34, and 35 weeks)
- Infants who qualified on seizures alone vs. not (2 levels)
- Infants who had a baseline temperature < 36.5 °C vs. not (2 levels)
- Infants who had a baseline temperature < 35.0 °C vs. not (2 levels)
- Infants who overshot the target and experienced esophageal temperatures < 32.0° C vs. not (an analysis using only babies in the cooling arm) (2 levels)
- Infants who had an abnormal cranial ultrasound after start of intervention but prior to final status will be compared to those who did not (2 levels).
- degree of NE (moderate versus severe)

6.5.9 Ad-hoc Analyses

Other ad hoc analyses, defined as analyses not detailed in this SAP, may be done in addition to those specified here as the primary and secondary papers are developed. These will be clearly labeled in any publication as ad hoc analyses motivated by our findings as we completed the planned data analyses.

6.6 Missing Data

No imputation for missing data will be done.

6.7 Statistical Software

SAS, R, JAGS, and/or Stan will be used to complete the analyses.

6.8 List of Displays for primary paper

The document MockTablesAppendix.docx contains table shells for specified summaries and analysis that will go into the primary paper.