Evaluation of Systemic Hypothermia Initiated After 6 Hours of Age in Infants ≥ 36 Weeks Gestation with Hypoxic-Ischemic Encephalopathy: A Bayesian Evaluation

Principal Investigator: Abbot Laptook

Final: July 31, 2007
Revised: November 19, 2007
Revised: January 03, 2008
Revised January 23, 2008
Revised August 5, 2008

Version: 1.2

TABLE OF CONTENTS

1.1 Hypothesis 1
2.1 Background 1
2.1.1 Preliminary Data 3
2.1.2 Study Rationale and Need to Conduct the Study in the NRN 4
3.1 Inclusion Criteria 4
4.1 Exclusion Criteria 6
5.1 Randomization and Stratification 6
6.1 Intervention 7
7.1 Discontinuation of Induced Hypothermia 7
8.1 Post Randomization Exclusion of Infants 7
9.1 Safety Monitoring of Control and Experimental Infants 8
10.1 Sedation/Analgesia/Anti-convulsants 8
11.1 Data Collection 8
12.1 Follow-up 8
13.1 Primary Outcome 9
14.1 Secondary Outcomes 9
15.1 Estimated Available Number of Patients 9
16.1 Data Analysis 11
16.1.1 Introduction to Bayesian methods 11
16.1.1.1 Differences Between Frequentist and Bayesian Statistics 11
16.1.1.2 Consideration of Prior Evidence 11
16.1.1.3 Resolving Controversy about Bayesian Methods 12
16.1.2 Analytic Approach for the Proposed Trial 13
17.1 Phases of the Study 18
18.1 Monitoring of Safety for the Trial 18
18.2 Interim Monitoring Plan 18
19.1 Budget and Justification 20
Appendix 1 22
References 29

Evaluation of Systemic Hypothermia Initiated After 6 Hours of Age in Infants ≥ 36 Weeks Gestation with Hypoxic-Ischemic Encephalopathy: A Bayesian Evaluation

Principal Investigator
Abbot Laptook

NICHD Network Subcommittee Members

Nehal Parikh, DO/Jon Tyson, MD	Joseph Lucke, PhD
Seetha Shankaran, MD	Claudia Pedroza, PhD
Ed Bell, MD	Abhik Das, PhD
Richard Ehrenkranz, MD	Rosemary Higgins, MD
Ronald Goldberg, MD	Angelita Hensman, RN
Namasivayam Ambalavanan, MD	Karen Osborne, RN

Objective: Evaluate whether induced hypothermia with body cooling initiated between 6-24 hours of age and continued for 96 hours in infants ≥ 36 weeks gestation with hypoxic-ischemic encephalopathy will reduce the incidence of death or disability at 18 months of age.

Study Design: Multicenter, randomized trial. The intervention of hypothermia will be unmasked.

Sample size: 168 infants

Eligibility criteria: Infants ≥ 36 weeks gestation with a pH (cord or neonatal) ≤ 7.0 or base deficit ≥ 16 mEq/L, **or** an acute perinatal event and either a 10 minute Apgar ≤ 5 or continued need for ventilation. All infants must have signs of encephalopathy at an age between 6-24 hours at the time of enrollment.

Study intervention: Infant will be randomized between 6-24 hours of age to either hypothermia or a non-cooled control group. Hypothermia will be achieved with whole body cooling to an esophageal temperature of 33.5°C using a Cincinnati Sub-Zero Hyper/Hypothermia device for 96 hours. Infants in the control group will have their core temperature using the esophageal site maintained at 37°C by appropriate servo control of the skin temperature. Cardio-respiratory, renal, metabolic, hematological and neurological status will be monitored along with esophageal, skin, and axilla temperatures during 96 hours of the intervention.

Primary outcome: The primary outcome will be death or moderate/severe disability at 18 months of age.

Sample size estimates: It is estimated that 168 infants (84 per group) can be enrolled within 3 years. This represents the largest number of infants anticipated to be enrolled in the longest time feasible for conducting the study

(3 years for enrollment and 1.5 years for follow-up). Hypoxia-ischemia severe enough to warrant a brain specific therapy such as therapeutic hypothermia is uncommon. Conventional levels of statistical precision typically used to test a hypothesis in clinical trials are unlikely to be obtained in rare conditions. Given the small sample size a Bayesian analysis will be used to provide a systematic analysis of the available data. The first year of the trial will be considered a pilot phase in view of the uncertainty of enrollment; if enrollment is excessively low, the trail will be discontinued. If enrollment is adequate the study will continue and additional infants will be recruited during the final two years of enrollment.

1.1 Hypothesis

The risk for death or disability among infants with perinatal hypoxia-ischemia and moderate or severe encephalopathy is reduced if systemic hypothermia (esophageal temperature of 33.5°C) is initiated after 6 hours of age and continued for 96 hours compared to infants with esophageal temperature maintained at 37.0°C.

2.1 Background

Perinatal hypoxia-ischemia represents the etiology for newborn encephalopathy in up to 30% of affected infants, and can result in death, CP, mental impairment and seizures.[1, 2] Management of infants with HIE has been limited to supportive intensive care without any brain oriented specific therapy. This approach is changing based upon laboratory and clinical trials of brain cooling. A small reduction in brain temperature of neonatal animals (as little as 2°C) favorably affects multiple processes involved in the pathogenesis of brain injury (energy state, excitotoxins, nitric oxide, apoptosis, etc.) and attenuates the extent of clinical and histological brain injury.[3] In two subsequent large randomized clinical trials using a cooling cap or cooling blanket to achieve brain hypothermia there was either a strong favorable direction of effect[4] or improvement in outcome with a cooling regimen.[5]

During the design of these two trials the time of initiation, depth, and duration of hypothermia were extrapolated from animal studies. The best studied parameter of hypothermia regimens among perinatal animals was the time to initiate cooling; brain hypothermia was effective in reducing brain injury when started at 1.5 hrs following ischemia, was less effective at 5.5 hrs and was not effective at 8.5 hrs following brain ischemia in fetal sheep.[6-8] Two subsequent reports have demonstrated that neuroprotection can be achieved when brain hypothermia is initiated beyond the window of 5.5 hours. Hypothermia initiated at 6 hours in 14 day rats undergoing hypoxia-ischemia and continued for 6 hours had lower infarct volume and better cerebral energy metabolite ratios compared to hypothermia of the same duration and initiated immediately, 2 or 4 hours following hypoxia-ischemia.[9] Hypothermia initiated at 12 hours in adult rats undergoing cerebral ischemia and continued for 5 hours mitigated the extent of necrosis in the lateral CA-1 hippocampus.[10] The majority of the literature however supports greater efficacy of hypothermia when initiated sooner rather than later following a hypoxic-ischemic event. Observations from adult animals suggest that extending the duration of hypothermia may offset delays in initiating cooling and result in neuroprotection.[11-13] It is unknown if the duration of the therapeutic window in human neonates is similar to fetal sheep and if a longer duration of hypothermia can widen the window. Although the latter remains in question, the pathogenesis of brain injury evolves over an extended interval of days to weeks. Evidence of a prolonged inflammatory response following hypoxia-ischemia supports this contention. IL-1β stimulates the synthesis of other cytokines, induces leukocyte infiltration, influences glial gene expression, and stimulates production of trophic factors.[14] Following hypoxia-ischemia in neonatal rats IL-1 activity increases transiently and reaches a peak at 6 hours.[15] The initial rise in IL-1β is followed by a secondary rise starting at 3 days and extends to 14 days after hypoxia-ischemia.[16]

Similar prolonged expression of other inflammatory mediators such as intercellular adhesion molecule 1 (ICAM-1) is detectable in adult animals. The latter is an endothelial ligand for the β-2 integrins on leukocytes. Post-ischemic influx of leukocytes into ischemic tissue may exacerbate injury and antibody to ICAM-1 reduces the extent of brain damage after middle cerebral artery occlusion (MCAO) in rats.[17] The temporal profile of ICAM-1 protein and mRNA expression after transient MCAO in adult rats peak at 12 hours and persists to one week of reperfusion.[18] Other potential endogenous repair mechanisms have a prolonged temporal profile after hypoxia-ischemia. For example the subventricular zone provides a progeny of reparative cells that appear to be stimulated by hypoxia-ischemia in 10 day old mice and is expanded for at least 2 weeks following hypoxia-ischemia.[19]

There also is an emerging body of animal investigations that demonstrate exacerbated neuronal injury with modest increases in brain temperature during ischemia (3°C).[20, 21] Even small, clinically relevant changes in temperature of only 1-2°C adversely affected post-ischemic neurological function and neuronal injury of adult dogs.[22] In adult gerbils a transient hyperthermia occurs during the early recirculation phase following brain ischemia and suppression of the hyperthermia by anesthetics attenuated injury to the hippocampus.[23] These findings were independent of the anesthetic effects. Of greater concern is that 3 hours of hyperthermia (39-40°C) initiated at 24 hours following brain ischemia increased ischemic neurons of the CA-1 sector by 2.5 fold compared to 38°C in adult rats.[24] Similar effects of elevated temperature are observed in neonatal animals. In 7 day rat pups an increase in brain temperature of 1-2°C during hypoxia-ischemia aggravated behavioral deficits and neuronal injury compared to normothermic animals.[25] In 10 day rat pups an increase in core body temperature (37.5°C compared to 36.0°C) for four hours immediately following hypoxia-ischemia increased the extent of neuronal injury.[26]

Hyperthermia acts through several mechanisms to worsen cerebral hypoxia-ischemia including enhanced release of neurotransmitters, exaggerated oxygen radical production, greater blood-brain barrier breakdown, impaired recovery of energy metabolism and protein synthesis and worsening of cyto-skeletal proteolysis.[27] The interest in elevated temperature of the fetus and newborn is reflected in associations reported between maternal fever and outcomes such as need for resuscitation, encephalopathy, seizures and cerebral palsy based on observational and case-control studies.[28-33] Recognition of the adverse effects of hyperthermia has prompted analysis of fever in adult stroke patients. A meta-analysis of 9 studies (3,790 patients) demonstrated an association between hyperthermia after stroke onset and an increase in morbidity and mortality (a test of heterogeneity was non-significant for mortality).[34] These studies have not rigorously excluded the possibility that larger strokes result in fever. Nevertheless, the effects of hyperthermia in adults is sufficiently concerning that it is recommended to vigorously minimize fever in patients with ischemia even if the extent of temperature elevation is considered minor or delayed in onset.[27] Only an interventional trial can determine if this association is causal.

2.1.1 Preliminary Data

Observations from the recently completed NICHD whole body hypothermia trial suggests that elevated body temperature is a frequent finding among infants with HIE cared for in the usual care group. Temperature regulation in the usual care group was initially servo control of abdominal wall skin temperature between 36.5-37.0°C with subsequent adjustments of the servo set point based on local practice of individual units (most commonly based on axillary temperatures). Each infant in the usual care group had esophageal temperatures recorded at 4 hour intervals during the 72 hour intervention period (total 19 values). Esophageal temperatures were not used in the management of infants in the usual care group.

The mean (± sd) esophageal temperature for infants in the usual care group was 37.2 ± 0.6°C over the 72 hour intervention. The distribution of all esophageal temperatures (n = 1839, 173 missing values) is listed in the table. Sixty-four percent of all esophageal temperature values in the usual care group were > 37°C. Of the 106 infants randomized to the usual care group, 4 had missing esophageal temperatures and 50 infants had a maximum esophageal temperature ≥ 38°C (the maximum temperature was 41.1°C). Even for the 52 infants with a maximum esophageal temperature < 38°C (n=970 values), the percent distribution for esophageal temperatures ≤ 36.5, 36.6-37, 37.1-37.5, and 37.6-38.0°C were 11.3, 31.1, 46.6, and 10.9%, respectively.

Esophageal Temperature (°C)	% of all Temperature Values
≤ 36	2.6
36.1 – 36.5	7.3
36.6 – 37.0	25.9
37.1 – 37.5	41.5
37.6 – 38.0	15.5
38.1 – 38.5	4.2
>38.5	3.0

The relationship between elevated temperature in the usual care group and the risk of an adverse outcome (death or disability) was examined in an observational study and presented at the May 2006 PAS meeting.[35] Logistic regressions were used to relate death or disability to measures of temperature for each infant adjusting for the level of encephalopathy, gender, race and gestational age. Separate regressions were created for measures of the highest and median temperature of esophageal or skin temperature and results were expressed as an odds ratio and 95% confidence interval. The measure of the highest temperature of each infant was represented by the average of the highest quartile of temperature collected during the 72 hour intervention period. The results indicated that an increase of only 1°C in the average of the highest quartile of skin or esophageal temperature was associated with a 3.6-4 fold increase in the odds of death or disability. The odds of death were increased 5.9 fold for each centigrade increase in the median esophageal temperature. There was no relation between the median skin temperature and outcome.

Given the observational study design, a casual inference between elevated temperature and outcome cannot be distinguished from elevated temperature

secondary to brain injury. However the results suggest that evaluation of neuroprotection associated with brain hypothermia or any other intervention should be compared to study patients in whom increases in temperature are avoided.

2.1.2 Study Rationale and Need to Conduct the Study in the NRN

The NRN trial of systemic hypothermia[5] is limited to infants that qualify in the first 6 hours following birth and it is unknown if systemic hypothermia is of benefit when initiated at a later age. The parameters of the cooling regimen were based on the best available animal data at the time of trial design and do not imply that these are optimal. There are three important scientific justifications to conduct this study: 1) animal data suggests that brain injury evolves over a prolonged time (days to weeks) following hypoxia-ischemia, 2) temperature modulation may have prominent effects on brain outcome even remote from the time of injury, and 3) prolongation of treatment with hypothermia may offset later initiation of the reduction in temperature. The relative importance of time of initiation and duration of hypothermia for the extent of neuroprotection is not known. The efficacy of brain cooling initiated at < 6 hours of age is based on controlled laboratory observations in the sheep fetus.[8] Given the uncertainty in determining the timing of a "hypoxic-ischemic event" for many infants in the prior Hypothermia trial, it is plausible that hypothermia was initiated more than 6 hours from the "event". If the results of the study are positive, more infants can be offered the therapy and outcomes can be improved. If the results do not demonstrate benefit, important information is provided that should limit the inappropriate use of this therapy.

This investigation will address a population of patients that could not previously be studied due to geographic considerations (inability to transport eligible infants within 6 hours), late recognition of encephalopathy, or progression of an encephalopathy beyond mild degrees of involvement.[36] Thus continued investigation of systemic hypothermia beyond its present use addresses gaps in knowledge concerning potential broader application of this therapy.

There are compelling reasons why this investigation should be conducted in the NRN. The NRN has an established follow-up program for the primary outcome with certified examiners trained to reliability, low attrition, and standardized assessments. Many of the present Network centers participated in the prior trial of systemic hypothermia and the subsequent free standing evaluation of the amplitude integrated EEG. This experience has provided an infrastructure for screening, identification and examination of infants, and familiarity with the intervention of systemic hypothermia. No other network is positioned to initiate this type of study with minimal resource investment as the NRN.

3.1 Inclusion Criteria

All infants with a gestational age ≥ 36 weeks will be screened for study entry if they are admitted to the NICU with an admitting diagnosis of neonatal depression,

perinatal asphyxia or encephalopathy. Infants will be evaluated in two sequential steps; evaluation by clinical and biochemical criteria (Step A) followed by a neurological exam (Step B). Details are as follows:

Step A: All infants between 6 and 24 hours of age will be evaluated for the following:

1. History of an acute perinatal event (abruptio placenta, cord prolapse, severe FHR abnormality, e.g., variable or late decelerations).
2. An Apgar score ≤ 5 at 10 minutes.
3. Continued need for ventilation initiated at birth and continued for at least 10 minutes.
4. Cord pH or first postnatal blood gas pH at ≤ 1 hour ≤ 7.0.
5. Base deficit on cord gas or first postnatal blood gas at ≤ 1 hour ≥ 16 mEq/L.

The above criteria are intended to screen for infants with a high probability of acute hemodynamic compromise around the time of birth. All of the above criteria do not need to be fulfilled in each patient. Two different pathways will be used as an indication of an acute event for the fetus/newborn. If a profound fetal acidemia is present category A1 (see below) is followed, and if either a blood gas is not available or the fetal acidemia on a blood gas is more modest, category A2 is followed.

IF BLOOD GAS IS AVAILABLE:	**IF BLOOD GAS IS NOT AVAILABLE OR pH between 7.01 and 7.15, OR BASE DEFICIT 10 to 15.9mEq/L**
A1	A2
Infant should have: (4 or 5 from above)	Infant should have: (1 and 2 or 3 from above)
	• Acute perinatal event and either
• Cord pH or first postnatal blood gas within 1 hour with pH ≤ 7.0	• An Apgar score ≤ 5 at 10 minutes
OR	**OR**
• Base deficit on cord gas or first postnatal blood gas within 1 hour at ≥ 16 mEq/L	• Continued need for ventilation initiated at birth and continued for at least 10 minutes

If the criteria in A1 or A2 are met, the infant qualifies for a neurological examination (Step B).

Step B. An abnormal neurological exam will be the presence of moderate or severe encephalopathy defined as seizures OR the presence of abnormality in at least 3 of the 6 categories in Table 1.

Table 1

Category	Moderate Encephalopathy	Severe Encephalopathy
1. Level of consciousness	Lethargic	Stupor/coma
2. Spontaneous activity	Decreased	No activity
3. Posture	Distal flexion	Decerebrate
4. Tone	a. Hypotonia (focal, general) b. Hypertonia	a. Flaccid b. Rigid
5. Primitive reflexes Suck Moro	 Weak Incomplete	 Absent Absent
6. Autonomic system Pupils Heart rate Respirations	 Constricted Bradycardia Periodic breathing	 Skew deviation or dilated, non-reactive to light Variable HR Apnea

The neurological examination will be performed by a physician examiner. To ensure compliance with the defined entry criteria and achieve consistency among examiners, all physician examiners will meet a standardized certification process. To facilitate the accuracy of the neurological examination, every attempt should be made to withhold the administration of medications that may alter the examination (e.g., versed, fentanyl etc) until after the exam is completed unless imperative for clinical care.

These criteria are identical to the completed Network Hypothermia trial[5] except for the time of entry (6-24 hrs vs < 6hrs of age). The amplitude integrated EEG will not be used as inclusion criteria since it remains uncertain whether the aEEG improves selection of infants at risk for death/disability compared to the above criteria. In addition use of the same entry criteria allows comparison with the completed Network Hypothermia trial. This study will recruit infants that qualify for brain hypothermia, but are not cooled either because they were not transferred to a center at less than 6 hours of age, the neurological status progressed to moderate/severe encephalopathy after 6 hours of age, their neurological findings were not recognized at < 6 hours of age, the equipment was not immediately available, or it was not feasible to examine the infant, obtain consent, and randomize before 6 hours.

4.1 Exclusion Criteria

1) Core body temperature (axilla, rectal) less than 34.0°C for greater than 1 hour, 2) presence of a known anomaly or chromosomal aberration, 3) birth weight <1800 gms, 4) infant in extremis, 5) refusal of parents or attending physician, 6) participation in conflicting clinical trial.

5.1 Randomization and Stratification

After informed consent is obtained, infants will be randomized to either an esophageal temperature of 37.0°C or 33.5°C for 96 hours. Enrolled infants will be stratified by age

of enrollment (≤ 12 and > 12 hours) and stage of encephalopathy (moderate or severe). It is anticipated that the majority of infants who will qualify for this study will fall within the 6-12 hour age range. A variable permuted, block size will be used.

6.1 Intervention

Infants randomized to cooling will be placed on a cooling/heating blanket which will be coupled to a Cincinnati Sub-Zero Hyper-Hypothermia Blanketrol System. Blankets can be positioned on radiant warmers, cribs, or isolettes. An esophageal temperature probe will be placed in the lower third of the esophagus and the probe will be interfaced with the Blanketrol System. The esophageal temperature will be controlled in the automatic control mode ("servo") at 33.5°C for 96 hours. No other heating mechanism will be used during this interval (all external heat sources must be off). At the completion of 96 hours, the control set point will be increased 0.5°C per hour until the esophageal temperature set point is 37.0°C for one hour. The remainder of the rewarming phase (5hr) will be completed with the Blanketrol in the "monitor only" mode and servo control of skin temperature with a radiant warmer will be used to maintain esophageal temperature of 37.0°C. Subsequent temperature control will be changed to the standards of the participating NICU.

Infants randomized to the non-cooled control group will have an esophageal temperature probe placed in the lower third of the esophagus and temperature monitored with either a Blanketrol system or a Mon-A-Therm dual input thermometer. Esophageal temperature will be controlled at 37.0°C by servo control of the abdominal skin temperature using an initial servo set point of 35.5°C. There will be an acceptable range of temperatures above and below 37.0°C beyond which a simple algorithm will be provided to respond to potential elevated temperatures. Following 108 hours of observations (to mirror the cooled group), the esophageal temperature probe will be removed and temperature control will be changed to the standards of the participating NICU.

7.1 Discontinuation of Induced Hypothermia

Infants in the cooled group will have hypothermia discontinued if any of the following occur: parents withdraw consent, Neonatologist withdraws consent or infant needs ECMO therapy. Discontinuation of hypothermia for a serious adverse event requiring therapy (one or more of the following: cardiac arrhythmia, persistent acidosis, major thrombosis or bleeding, skin breakdown or equipment malfunction) is at the discretion of the attending physician after consultation with the study/site PI.

8.1 Post Randomization Exclusion of Infants

The study is designed as intent-to-treat, and therefore infants will not be excluded after randomization.

9.1 Safety Monitoring of Control and Experimental Infants

a. Skin, esophageal, axilla, and servo set point temperature will be monitored hourly until 12 hours and every 4 hours thereafter during the 96 hour intervention interval and subsequent 12 hours (total of 108 hr). For infants undergoing hypothermia, temperature will be recorded every 15 minutes for the first 3 hours of cooling, then hourly until 12 hours and then every 4 hours thereafter.
b. Metabolic status: serum electrolytes will be monitored as per clinical routine.
c. Respiratory status: blood gases will be monitored as per clinical routine.
d. Cardiovascular: heart rate, blood pressure and use of inotropic agents will be recorded at baseline and every 4 hours for 96 hours.
e. Renal status: urine output and body weight will be recorded daily during the intervention interval. Serum BUN and creatinine will be monitored as per clinical routine.
f. Neurological status: at baseline, after rewarming, and at discharge (performed by certified examiner).
g. Hematological: PT/PTT will be acquired only if bleeding is suspected based upon clinical symptoms or an unexplained fall in hematocrit by more than 10%. Complete blood counts will be monitored as per clinical routine.
h. Infectious Disease: Results of blood cultures will be recorded.

10.1 Sedation/Analgesia/Anti-convulsants

The use of sedative, hypnotic and analgesic agents, and anti-convulsants will be at the discretion of the Attending physician.

11.1 Data Collection

Data will be collected on maternal labor events (including presumed chorioamnionitis and antibiotic use), mode of delivery, infant characteristics and demographics, delivery room events, Sarnat stage at randomization, and 96 hours of age, occurrence of seizures, evidence of other organ dysfunction, use of anti-convulsants and sedative-hypnotic agents, results of CNS imaging and EEG studies and neurological exam at discharge. Determination of the stage of encephalopathy will be based on a modified Sarnat stage by scoring the presence of moderate or severe abnormalities in 6 categories. The number of moderate or severe signs determines the extent of encephalopathy and if signs are equally distributed the designation of moderate or severe encephalopathy is based on the level of consciousness. Prior to the intervention temperature measurements will be recorded for both in-born and transported infants. Esophageal, skin, axillary, blanket and control set point temperature will be recorded at 15 min intervals during the first 3 hours of the intervention, hourly for 9 hours and then at 4 hour intervals until 108 hours. Orders for "do not resuscitate" and withdrawal of support will be recorded.

12.1 Follow-up

All surviving infants will be followed in the Neonatal Research Network follow-up program with a compliance rate maintained at 95%. Tracking information will be

recorded at the time of discharge from the NICU. An attempt will be made to obtain an autopsy in case of death occurring prior to and following NICU discharge.

Growth parameters, a neurological examination, psychometric testing, and vision and audiometric evaluations will be performed as part of the follow-up evaluations. In addition, the family's socio-economic and educational status will be assessed. Infants will be tracked and undergo follow-up at Network centers with evaluations at 18-22 months of age by personnel trained to reliability and blinded to treatment assignment group. If an infant is not evaluated at the 18-22 month clinic visit because of acute illness, behavior problems, or "other" reasons, appointments will be re-scheduled until the evaluation is complete.

13.1 Primary Outcome

The primary outcome will be death or disability (either moderate or severe in extent). Severe disability will be defined by any of the following: a Bayley III cognitive score < 70, Gross Motor Functional (GMF) Level of 3-5, blindness or profound hearing loss requiring amplification but still unable to follow commands/communicate. Moderate disability will be defined as a Bayley III cognitive score between 70-84 and either a GMF level of 2, a seizure disorder, or a hearing deficit.

Infants without the primary outcome will be categorized as normal or mildly impaired. Normal will be defined by a Bayley III cognitive score ≥ 85, and absence of any neurosensory deficits. Mild impairment will be defined by either a Bayley III cognitive score of 70-84 alone, or a Bayley III cognitive score ≥ 85 and any of the following: presence of a GMF level 1 or 2, seizure disorder or hearing loss.

14.1 Secondary Outcomes

- Number of deaths in the NICU and following discharge
- Number of infants with moderate and severe disability
- Number of infants with mild, moderate and severe disability
- Number of infants with any disability based on level of encephalopathy at randomization
- Number of infants with non-CNS organ system dysfunction
- Number of infants with a DNR order
- Number of infants with a DNR order and support is withdrawn
- Number of infants with a DNR order and either die or survive
- Number of infants with neonatal seizures, with and without EEG abnormalities

15.1 Estimated Available Number of Patients

Estimates of the number of available patients for conducting this trial within the Network include 1) infants who do not develop seizures or moderate/severe HIE until after 6 hours age and 2) infants who cannot be randomized by 6 hours of age (late time of referral, unavailability of study personnel).

An estimate of the number of patients in the first category can be extrapolated from a study performed at UT-Southwestern comparing the Amplitude Integrated EEG and the neurological examination.[36] Entry criteria for this study was similar to the inclusion criteria stated above and infants (n=50) were categorized in the first 6 hours as normal or demonstrating features of encephalopathy (modified Sarnat stages I, II or III). There were 15 normal, 3 stage I, 17 stage II, and 2 stage III; 13 infants had features of both Sarnat stage I and II and could not be categorized definitively as either stage; 3 of these 13 progressed to a stage II with an abnormal short term outcome. Based on these data 9.7% (3/31) of infants who would not qualify for entry into the Network Hypothermia study during the first 6 hours (normal, Stage I, and those with features of both Stage I and II) may progress after 6 hours of age. If this is extrapolated to the screened (n=798) but not eligible infants (798-238 = 560) in the Network Hypothermia trial (conducted over 35 months) there are 54 potential candidate infants for a study conducted over 35 months.

An estimate of the number of patients in the second category can be extrapolated from the screening log and eligibility forms (IHO1 and IHO2) from the prior NRN systemic hypothermia trial.[5] Of the 798 infants in the screening log 78 infants (9.7%) were excluded based on unable to randomize by 6 hours of age. If this is extrapolated to the screened but not eligible infants in the NRN Hypothermia trial (n=560) another 54 infants may be eligible (35 month study). The latter figure should be considered an underestimate since some infants that may qualify were probably never referred to Network centers for the Hypothermia study given the age constraints (< 6 hours of age). If it is estimated conservatively that one infant per center per year is recruited based on the time of entry criteria (6-24 hours) beyond that of the screening/eligibility forms of the prior study, another 47 infants may be eligible for a 35 month study conducted in 16 centers.

These two sources of patients provide an estimate of 155 patients over 35 months or 160 patients extrapolated to a 3 year interval. Given the limited number of patients within the Network a randomized study could be done with the explicit purpose to obtain the most precise and unbiased estimate of the relative risk for death or disability that is reasonably feasible. The longest trial that has been considered reasonably feasible in the Network has been approximately five years. We propose to enroll 168 infants in a trial that would require approximately 3 years for enrollment and two years for follow-up. This provides 160 infants for analysis assuming a lost to follow-up of no greater than 5%. Conducting a longer trial (or incurring the additional expense, effort, and uncertainty inherent in involving Centers outside the Network) is unlikely to be acceptable to the Steering Committee.

This approach to sample size—assessment of the largest number of patients that is reasonably feasible in order to obtain the most precise and unbiased estimate of treatment effects—differs from the conventional (frequentist) approach to determining sample size. Frequentist statisticians ordinarily recommend against conducting a randomized trial when it is not feasible to achieve conventional sample size estimates. However, the alternative in this circumstance is to conduct no trial and to

allow the use of hypothermia beyond 6 hours to creep into clinical practice based on anecdotal experience or at best observational studies. As Schulz and Grimes have emphasized, assertions that trials should not be conducted "unless an arbitrarily defined level of statistical power can be assured make no sense if the alternative is acquiescence in ignorance of the effects of healthcare interventions...Unbiased trials with no results trump no results at all.[37]

Based on the largest feasible randomized trial, conventional statistical analyses may be used to provide an unbiased estimate of treatment effect with a 95% confidence interval. As discussed below, we propose Bayesian methods of analysis to indicate the probability of a clinically important effect.

16.1 Data Analysis

The use of Bayesian methods in clinical trials is rapidly increasing, and these methods have particular advantages for small or medium size trials.[38] For these reasons, a brief description of Bayesian methods is provided below.

16.1.1 Introduction to Bayesian methods (A technical summary of the Bayesian statistical model for this proposal is included as an Appendix 1)

16.1.1.1 Differences Between Frequentist and Bayesian Statistics

Conventional frequentist inference uses objective probabilities for the assessing the relative frequency that a statistical procedure is correct in an infinite sequence of replications. Bayesian inference uses subjective probabilities as logic for maintaining coherent beliefs. Frequentist inference focuses on the probability that the observed or more extreme data would be obtained assuming that the null hypothesis is true. Bayesian inference focuses on the probability that a hypothesis is true given the observed data. For example, suppose a clinical trial identified an excess of infants with death among control infants compared to infants receiving the intervention under investigation. Frequentist statistics would address the question: What is the likelihood that this excess or a larger difference between groups would have occurred if the null hypothesis—e.g., that treatment has no effect on mortality—is correct (i.e., Pr (observed or more extreme data $| H_0$)? Bayesian statistics addresses a fundamentally different question that addresses more directly what physicians want to learn from the trial: What is the probability that the intervention has no effect on mortality (or conversely, the probability that it does), given the data obtained in the trial (i.e.,Pr($H_0 |$ data).

16.1.1.2 Consideration of Prior Evidence

Any Bayesian analysis consists of combining prior evidence (represented as a *prior* distribution) with data (represented as a likelihood) obtained from the current study. The result is a *posterior distribution* from which all inference is derived. Thus a Bayesian statistician must quantify prior evidence in the form of a prior distribution.

This is in contrast with the frequentist approach where the analysis of a clinical study does not involve consideration of the evidence from other relevant studies. The frequentist statistician leaves it to the reader to consider the results from other studies in interpreting the current study results and reaching a conclusion or making a treatment recommendation. In contrast, Bayesian analyses involve formal consideration of prior evidence (before seeing the results of the trial) to estimate the posterior likelihood of the null hypothesis (after reviewing trial results). As an example, the posterior probability that an intervention reduced mortality as determined after completing a new trial may be assessed in a cumulative meta-analysis in which the prior probability is based on the results of all prior trials of the intervention. This approach to assessing treatment effect considering all prior and current evidence has been adopted by clinicians who use cumulative meta-analyses. Bayesian inference uses subjective probabilities with objective learning; classical inference uses objective probabilities with subjective learning.

16.1.1.3 Resolving Controversy about Bayesian Methods

Bayesian analyses have been criticized for the possibility of producing posterior probabilities that may hinge on excessively dogmatic or highly idiosyncratic beliefs about the phenomena under investigation or on widely varying beliefs among different investigators. This criticism has been largely addressed by the following kinds of responses from Bayesian statisticians:

1. The investigator's prior is open to public scrutiny, forcing him to make explicit his beliefs, justifiable or speculative, about the phenomenon under study, and thereby promoting honesty in the analysis.[39]
2. Accordingly, one's prior probability should be justified by the relevant evidence over speculation. Wherever possible, the prior probability should be based on standardized assessments of high quality studies, e.g. effects obtained from a well-performed meta-analysis of randomized trials.
3. One's prior may be *tempered* to take into account the beliefs of the scientific community and specialists in the field.[40]
4. There exist *reference* priors that have little influence on the data.[41] These may be used to obtain standard analyses relatively free of prior beliefs.
5. Bayesian analyses can be conducted using skeptical values for prior probabilities to determine whether the study results are sufficiently compelling to be persuasive even to skeptics.

Whether or not large trials have been performed, Bayesian statistics provides a formal quantitative method to assess the range of treatment effect compatible with the best available evidence and to estimate the probability of a clinically important benefit, considering all relevant prior evidence as viewed from the perspective of skeptics, enthusiasts, or physicians in equipoise. For these and other reasons, including the development of Bayesian software, the use of Bayesian methods in the design, monitoring, and analysis of clinical trials is progressively increasing For

example, the FDA recently encouraged the use of adaptive study designs based on Bayesian techniques in Phase I and Phase II trials.[38] Whatever the phase or sample size for a trial, Bayesian approaches provide a method to assess the range of treatment effect compatible with the data and to estimate the probability of a clinically important effect.

16.1.2 Analytic Approach for the Proposed Trial

The following is presented to illustrate the application of Bayesian methods to the kind of results that may be obtained in the proposed trial (n =160). Treatment effect is expressed as relative risk for death or impairment in the hypothermia group relative to that in the control group. Bayesian analyses are discussed below for circumstances in which the findings are compatible with moderate benefit (relative risk = 0.72), major benefit (relative risk = 0.64), no benefit (RR = 1.00) or harm (RR =1.10).

1. A relative risk of 0.72 is identified in the trial.[1] This relative risk would be identical to that in the prior trial, as could occur if the increased effectiveness due to a longer treatment period completely offset a reduction in effectiveness due to a delay in initiating therapy.

As indicated above, the posterior probability computed in a Bayesian analysis would depend on the prior probability, and the prior probability will vary depending on whether the viewpoint adopted is neutral (in equipoise), skeptical, or enthusiastic. The neutral prior would be that the relative risk would be 1.00. A skeptical prior would be that overall effect of hypothermia is actually harmful in this setting with a relative risk of 1.10. An enthusiastic prior would be that the relative risk would be 0.72 as in the prior trial.[5] To reflect the range of values that would be considered plausible from each perspective, the 95% credible intervals are.73 – 1.36 for the neutral prior,.83 – 1.48 for the skeptical prior, and.48 – 1.03 for the enthusiastic prior. Shown below are three figures indicating the values for the posterior probability that would be obtained for each of the 3 perspectives.[42] These figures correspond to Figures 1, 2 and 3 of reference 42 by Lilford and colleagues.

Figure 1. Prior belief is neutral (in Equipoise).

[1]This relative risk would be obtained if death or impairment occurred in 36 of 80 hypothermia infants and 50 of 80 control infants.

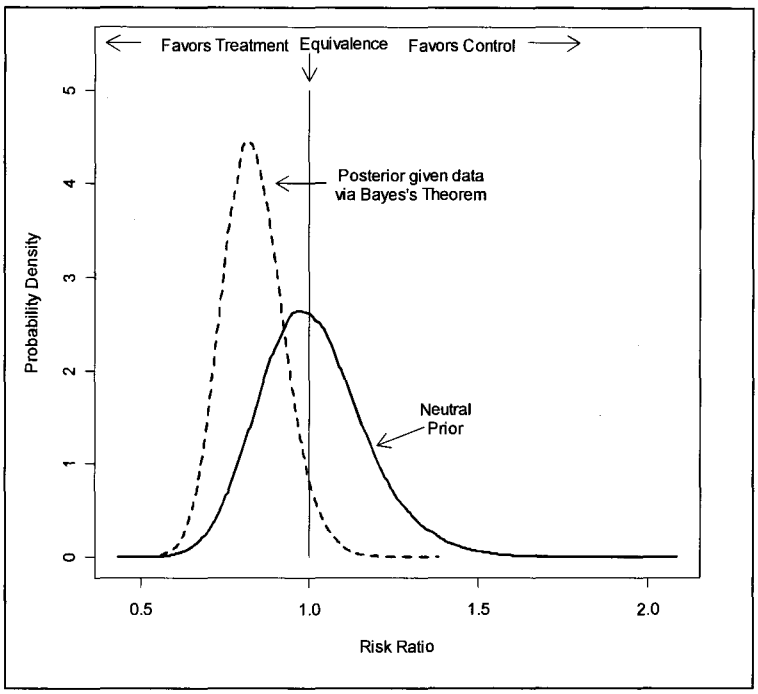

Figure 2. Prior belief is skeptical.

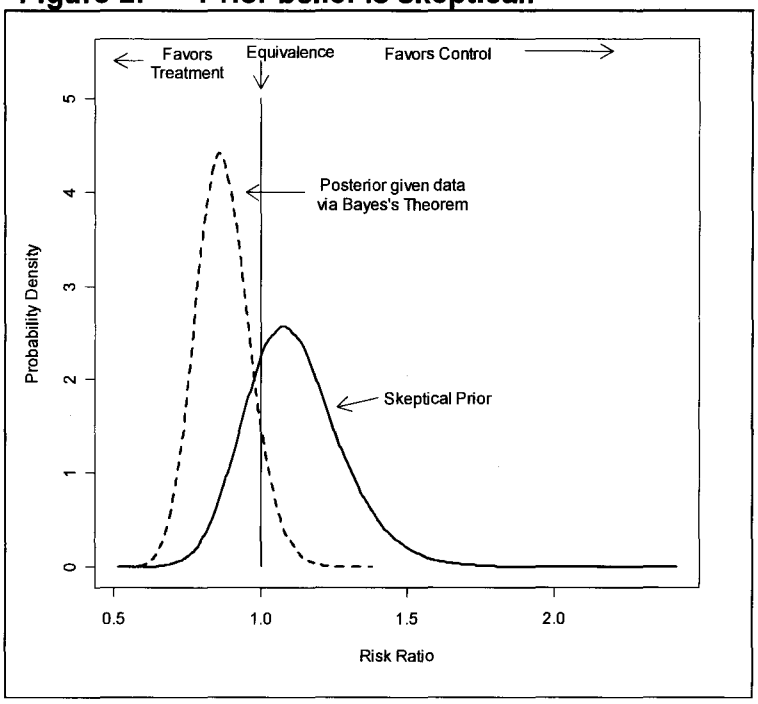

Figure 3. Prior belief is enthusiastic.

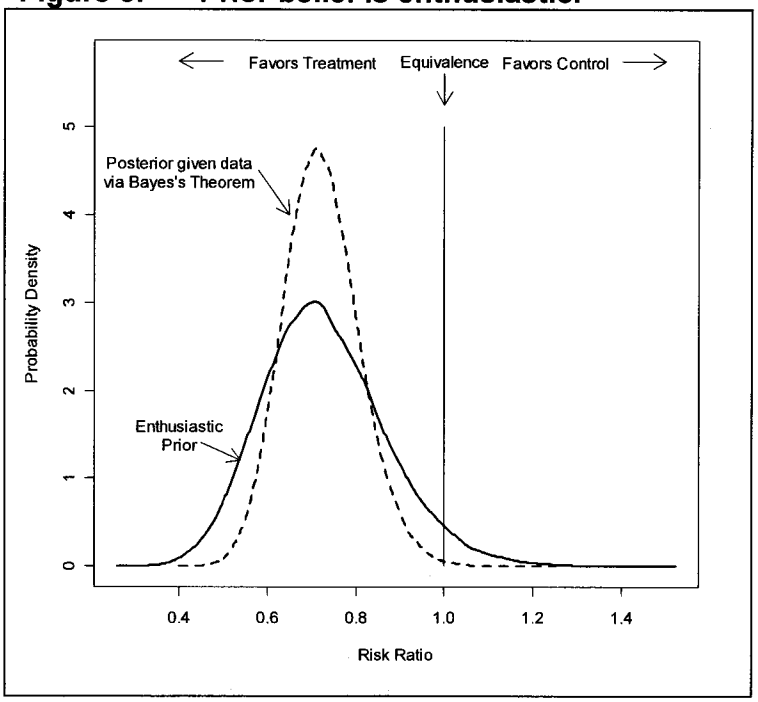

Table 2 below reflects the findings shown in the figures and provides the posterior probability of benefit reducing death or impairment exceeding 0%, 10%, or 20% (i.e, a true relative risk <1.0, <0.9, or <0.80).

Table 2

	Posterior Probability of Benefit					
	>0% reduction in death or impairment		**>10% reduction in death or impairment**		**>20% reduction in death or impairment**	
Perspective	Prior	Posterior	Prior	Posterior	Prior	Posterior
Neutral	**.50**	**.96**	**.25**	**.78**	**.07**	**.37**
Skeptical	**.25**	**.91**	**.08**	**.64**	**.01**	**.22**
Enthusiastic	**.96**	**.99**	**.89**	**.97**	**.71**	**.82**

Thus, if a neutral perspective is adopted in analyzing the trial, the posterior probability of at least some benefit would be **96%**, a 10% reduction in death or impairment would be **78%**, and the posterior probability of a 20% reduction would be **37%**. Even if a skeptical perspective is adopted, the corresponding posterior probability of at least some benefit would be **91%**; the value for a 10% benefit would be **64%**, and the value for a 20% probability would be **22%**.

2. A relative risk of 0.64 is identified.[2] Such a benefit could occur if the delay in treatment beyond 6 hours does not appreciably affect the benefits of therapy and/or the increase in duration of therapy has strong beneficial benefits. For the sake of brevity, figures are not displayed. However, as shown in Table 3 below, a high probability is obtained for all levels of benefit, even if a skeptical perspective is adopted.

Table 3

	Posterior Probability of Benefit					
	>0% reduction in death or impairment		**>10% reduction in death or impairment**		**>20% reduction in death or impairment**	
Perspective	Prior	Posterior	Prior	Posterior	Prior	Posterior
Neutral	**.50**	**.99**	**.25**	**.90**	**.07**	**.59**
Skeptical	**.25**	**.97**	**.08**	**.81**	**.01**	**.41**
Enthusiastic	**.96**	**.99**	**.89**	**.99**	**.71**	**.93**

Thus, if a neutral perspective is adopted in analyzing the trial, the posterior probability of at least some benefit would be **99%**; a10% reduction in death or impairment would be **90%**, and the posterior probability of a 20% reduction would be **59%**. Even if a skeptical perspective is adopted, the corresponding posterior probability of at least some benefit would be **97%**; the value for a 10% benefit would be **81%**, and the value for a 20% probability would be **41%**.

3. A relative risk of 1.0 is identified. Such a value could be obtained if there was no benefit from hypothermia initiated after 6 hours age or if the harm from extended use completely offset the benefit. As shown in Table 4 below, the value for the posterior probability would not justify administration of hypothermia even if an enthusiastic prior were adopted.

Table 4

	Posterior Probability of Benefit					
	>0% reduction in death or impairment		**>10% reduction in death or impairment**		**>20% reduction in death or impairment**	
Perspective	Prior	Posterior	Prior	Posterior	Prior	Posterior
Neutral	**.50**	**.50**	**.25**	**.14**	**.07**	**.01**
Skeptical	**.25**	**.34**	**.08**	**.06**	**.01**	**.003**
Enthusiastic	**.96**	**.87**	**.89**	**.54**	**.71**	**.15**

[2]This relative risk would be obtained if death or impairment occurred in 32 of 80 infants in the hypothermia group and 50 of 80 infants in the control group.

Thus, if a neutral perspective is adopted in analyzing the trial, the posterior probability of at least some benefit is **50%**; 10% reduction in death or impairment would be **14%**, and the posterior probability of a 20% reduction would be **1%**. If a skeptical perspective is adopted, the corresponding posterior probability of at least some benefit would be **34%**; the value for a 10% benefit would be **6%**, and the value for a 20% probability would be **0.3%**. If an enthusiastic perspective is adopted, the corresponding posterior probability of at least some benefit is **87%**; 10% reduction in death or impairment would be **54%** and the value for a 20% probability would be **15%**.

1. A relative risk of 1.10 is identified. [3] Such a value could be obtained if the harm from extended use was greater than any benefit.

Thus, if a neutral perspective is adopted in analyzing the trial, the posterior probability of any benefit is **26%**; a 10% reduction in death or impairment would be **4%**, and the posterior probability of a 20% reduction would be **0.1%**. If a skeptical perspective is adopted, the corresponding posterior probability of at least some benefit would be **14%**; the value for a 10% would be **1%**, and the value for a 20% probability would be **0.02%**. If an enthusiastic perspective is adopted, the corresponding posterior probability of at least some benefit would be **69%**; the corresponding posterior probability of a 10% reduction in death or impairment would be **29%** and the value for a 20% probability would be **4%**.

One might view the above results with concern that a clinician with an enthusiastic viewpoint may interpret the posterior probabilities as an indication to adopt the treatment even if the relative risk for a reduction in death or impairment is 1.0 or 1.1. However the viewpoint of the enthusiast is not advocated and is presented for sake of completeness. With a relative risk of 1.0, the enthusiast may want to adopt the treatment given an 87% probability of some reduction in death or impairment; however the likelihood of finding a > 10% benefit is not much better than 50:50 and the investment of time, cost and the possibility of harm in using the treatment may temper even the enthusiast. With a relative risk of 1.1 there is a dramatic reduction in the enthusiasts' belief that the treatment would yield some benefit; before observing the data the enthusiast was 96% sure of at least some effect and after performing the trial he/she is only 69% sure. The results have considerable impact even on the enthusiast. If one were to adopt the treatment that has a relative risk of 1.1, they would be doing so in the face of a 31% (1 – 0.69) chance of no benefit or harm, and a 71% chance of not having a 10% reduction in death or disability.

Results of the Bayesian analysis will be adjusted for center and gender.[43, 44]

[3]This relative risk would be identified if death or impairment occurred in 55 of 80 infants in the hypothermia group and 50 in the control group.

For comparison, the frequentist required sample size for a trial with an alpha error of 0.05, 80% power and a relative risk of 0.64, 0.72, 0.9 and 1.1 are 80, 134, 1054 and 991 subjects per group, respectively.

17.1 Phases of the Study

The study will be initiated as a pilot for 12 months given the uncertainty about the number of patients available for enrollment. A 12 month interval should provide a long enough time to determine the ability to recruit for this trial given the variability in start time among centers (IRB issues, in-services, education of referring centers etc). A commitment to complete the entire study will not be made until the end of the pilot period. This decision will be made by the DSMC along with review of the data for any unexpected toxicity associated with a longer interval of systemic hypothermia. Outcomes (death/disability) for infants in the pilot phase are not required for continuation of the study. All infants enrolled in the pilot phase will be used as patients in the main trial since the protocol is identical.

18.1 Monitoring of Safety for the Trial

- The protocol will be reviewed by the Institutional Review Board of each participating institution.
- Adverse events will be reported on the MedWatch form to the Data Center of the NICHD Neonatal Network, Research Triangle Institute (RTI), Chapel Hill, North Carolina. Adverse events will be compared between the two groups using the approach outlined in section 18.2. RTI will be responsible for reporting adverse events to the Data Safety Monitoring Committee (DSMC) of the Network.
- All protocol deviations will be monitored by RTI.
- RTI will prepare reports for presentation to the DSMC at periodic intervals.
- DSMC will be responsible for monitoring the safety and efficacy of the trial.

18.2 Interim Monitoring Plan

The intervention for the Late Hypothermia Study for HIE will be conducted during the first 4 days of life for the study subjects, while the primary outcome of death or disability will be assessed at 18-22 months of age. Thus, similar to the original Neonatal Research Network (NRN) Hypothermia trial, interim monitoring for this trial will be based primarily on monitoring adverse events for 124 hours from baseline (time of insertion of esophageal probe), with no explicit plans for monitoring treatment efficacy, given that the primary outcome is only available after 18-22 months. Adverse events monitored will include the following, plus a composite (any of the listed events), on which statistical comparisons between the treatment groups will be based:

- Cardiac arrhythmia
- Persistent acidosis
- Thrombosis

- Bleeding
- Alteration of skin integrity
- Death
- Any of the above.

In addition, we will also collect data/compare the rates of seizures during the 12 hours of rewarming. We propose to monitor the difference between the two treatment groups with regard to the incidence of these adverse events after every 20 babies are enrolled in the trial for 124 hours.

In keeping with the Bayesian nature of the trial design, we will use a Bayesian interim monitoring approach to determine whether systematic differences are emergent among the two treatment groups in the incidence of any of the adverse events listed above. Specifically, after every 20 babies are enrolled in the trial for 124 hours, we will calculate the posterior probability that one group has higher incidence of any adverse events than another. In statistical terms, the treatment will be considered harmful (i.e., the DSMC may consider termination of the trial) if for a pre-specified threshold η, the posterior probability of treatment harm (in terms of the above adverse events) is greater than η; in other words if the predictive probability $P(\theta > 1 \mid X) > \eta$, where θ denotes the relative risk favoring the treatment group and X is the data available. In order to allow for a liberal safety monitoring regime to ensure patient safety, we propose to take η = 80%. However, we will also present the 95% credible interval for θ, as well as the entire posterior distribution for θ in a graphical fashion to the DSMC so that they have a full appreciation for the range of possible values of θ.

The choice of a prior distribution is essential and controversial in Bayesian analyses. Thus, we can compute the posterior probability of treatment harm and the posterior distribution of the associated relative risk under two sets of priors – (a) a non-informative prior that does not assume any substantive prior information about treatment differences in terms of the adverse events listed above, and (b) an informative prior based on treatment differences for these adverse events observed in the original NRN Hypothermia trial. However, we realize that the standard Bayesian approach of running the analysis with different priors provides only a partial solution, as the results on stopping are often inconclusive, especially when few data have accumulated. Instead of worrying about the selection of a single correct prior, we will also explore the use of a robust Bayesian approach which replaces a prior distribution with a class of priors and calculates the corresponding posterior probabilities for decision making. [45]

19.1 Budget and Justification

Table 6 shows an estimated budget is provided for the entire trial assuming enrollment of 168 patients with outcomes available for 95% of the patients (160 patients).

	Cost ($ per patient, device or center)	Number of patients, device or centers	Total cost - $
Research time	1,500/patient	168	252,000
Medical Supplies	60/patient	168	10,080
Follow up	1,200/patient	160	192,000
Blanketrol II machine for new centers	4,100/device estimated	4 sites, 6 devices	24,600
Training meeting	500	16	8,000
Start up costs	1500/center	16	24,000
Total			***510,680***

Research time: Costs will cover time to screen and determine eligibility of patients, data collection, initiating and monitoring of the cooling intervention, and transmission of all data items. There may be further discussion needed regarding dedicated funds for being on-call.

Medical supplies: Costs will cover supplies for the Cincinnati Sub-Zero Blanketrol including temperature probes, thermal blankets, and temperature probe adaptors.

Follow-up: Costs will cover tracking infants, incentives to participate in Follow-up and performance of follow-up at Network sites. The higher costs of follow-up for this study are based upon a) this group of patients are not routinely followed by the Network, b) poor outcomes may be common and require higher incentives for participation, and c) the absence of a brain specific treatment for infants in this age group (after 6 hours of age) may result in infants transported from very far distances and require higher follow-up costs.

Blanketrol devices: Each new network center will require Cincinnati Sub-Zero Hyperthermia-Hypothermia Blanketrol II devices and the costs listed are based on a price estimate from the company. The University of Utah is planning on enrolling at 3 different hospitals. Cincinnati Sub-Zero now makes a Blanketrol III device and has given a price quote of $7,350 per unit. The relative merits of this device compared to the Blanketrol II are unclear at present but may deserve discussion.

Training meeting: The study PI from each Network site will be required to attend one training session in conjunction with the Steering Committee prior to initiation of the trial. Funds are required to cover the cost of travel and lodging assuming this would occur following a Steering Committee meeting.

Start-up costs: Funds are required for the time to train personnel and implement the study within sites.

Revised August 5, 2008
Revised January 23, 2008
Revised January 3, 2008
Revised November 19, 2007
Final July 31, 2007

Appendix 1

Bayesian Statistical Model for Hypothermia Study

Prepared by Joseph F. Lucke
Center for Clinical Research and Evidence-based Medicine
UTHSC-Houston

1 Introduction

The following is a technical report explicating the proposed Bayesian statistical model for *Evaluation of Systemic Hypothermia Initiated after 6 Hours of Age in Infants Greater than 36 Weeks Gestation with Hypoxic-Ischemic Encephalopathy* with Dr. A. Laptook, Principal Investigator.

2 Primary Outcome

The primary outcome is the composite of death or survival moderate to severe impairment versus the composite survival with mild to no impairment at 18–22 months.

3 Treatment and Stratification

The treatment will be hypothermia versus standard care. Prognostic stratification variables will be *age* at enrollment (age $\leq$ 12 hrs versus age $>$ 12 hrs), and *stage* of encephalopathy (moderate versus severe), and their interaction. Infants will be randomly assigned to treatment within stratification by *age* and *stage*. Infants will be stratified by, but not randomized within, site.

4 Model for Data Generation Process

A Bayesian model comprises three components: a model for the *data generating process* or *likelihood*, the *prior* distribution of beliefs regarding the parameters before the data are observed, and the *posterior* distribution of beliefs regarding the parameters after the data are observed. The first two components together with the data determine the last. Here the data generating process is presented.

Let N denote the number of HIE infants in the study, and let S denote the number of sites from which observations will be collected. (Currently N is anticipated to be about 160 and S is 16) Let $y_i \in \{0, 1\}$ denote the binary outcome:

$$y_i = \begin{cases} 0 & \text{for survival with no to mild disability} \\ 1 & \text{for death or survival with moderate to severe disability} \end{cases}$$

for infant i, $i = 1, \ldots, N$. Also for infant i, let s_i denote the site number, z_i the treatment, and $\mathbf{x}_i$ the p-vector ($p = 3$) of prognostic covariates, where

$$s_i \in \{1, \ldots, S\},$$

$$z_i = \begin{cases} 0 & \text{for standard care} \\ 1 & \text{for hypothermia treatment;} \end{cases}$$

$$x_{1i} = \begin{cases} 0 & \text{for age} \leq 12 \text{ hrs} \\ 1 & \text{for age} > 12 \text{ hrs;} \end{cases}$$

$$x_{2i} = \begin{cases} 0 & \text{for moderate encephalopathy} \\ 1 & \text{for severe encephalopathy;} \end{cases}$$

and

$$x_{3i} = x_{1i}x_{2i} = \begin{cases} 0 & \text{for age} \leq 12 \text{ hrs or moderate encephalopathy} \\ 1 & \text{for age} > 12 \text{ hrs and severe encephalopathy.} \end{cases}$$

We make the standard assumption that a treatment for any particular infant has no effect on the outcome of any other infant and that the process of measuring an outcome has not effect on the outcome (Gelman, Carlin, Stern, & Rubin, 2004). We also assume the infants' outcomes are mutually independent. Therefore, the outcome y_i follows a Bernoulli distribution whose probability parameter π is a function of the site, treatment, and predictors associated with the infant:

$$y_i \mid (s_i, z_i, \mathbf{x}_i) \sim \text{Bernoulli}\left[\pi\left(s_i, z_i, \mathbf{x}_i\right)\right].$$

The effects of site, treatment, and covariates are modeled as a logistic regression:

$$\text{logit}\left[\pi\left(s_i, z_i, x_i\right)\right] = \alpha_{s_i} + \theta z_i + \mathbf{x}_i'\boldsymbol{\beta} \tag{1}$$

The parameter θ represents the (randomized) treatment effect, expressed as log-odds, with $\theta < 0$ favoring treatment. The treatment is assumed to be fixed and modified neither by site nor by the covariates.

The parameter $\boldsymbol{\beta}$ represents the prognostic covariate effects. These effects are assumed to be fixed and modified neither by site nor by treatment. These covariates are included to increase the precision of the treatment effect.

The influence of the sites on outcomes can be modeled in one of three different ways: complete pooling, no pooling, or partial pooling (Gelman & Hill, 2007). *Complete pooling* regards the sites as homogeneous, having no differential effects, but serving merely as independent replications within the overall study. In this case, all the site effects are presumed equal so that $\alpha_s = \alpha$ for all s. *No pooling* regards the sites as separate studies such that the outcomes from any one site provide no information regarding those from any other site. Each site would thus require its own parameter α_s. *Partial pooling* regards the sites as having heterogeneous effects, but without any additional information as to which site would exhibit which effect.

Partial pooling is assumed here and is modeled as *exchangeability* among sites (Gelman et al., 2004). From the exchangeability assumption, the effect of any site can be modeled as arising from a super-population of site effects. Here we assume

$$\alpha_s \sim \text{normal}(\mu_\alpha, \sigma_\alpha^2). \tag{2}$$

The site effects are assumed to be modified neither by treatment nor by the covariates.

Thus, the model for the data generating process, i.e. the likelihood, is a *multilevel logistic regression with a varying intercept* (Gelman & Hill, 2007). This model is a extension of the model in the proposal to include sites and prognostic covariates.

5 Prior Densities

The second component is the *prior* distribution or density regarding the parameters. In general, th elicitation of prior densities can be difficult (Garthwaite, Kadane, & O'Hagan, 2005), but in our case, the elicitation of priors is considerably simplified. Except for the treatment parameter θ, the priors for the site and covariate effects will be *diffuse*, *vague*, or *weakly informative* so that they have little influence on the observations.

The priors for μ_α and σ_α^2 in Equation 2

$$\mu_\alpha \sim \text{normal}(0, 10^2);$$
$$\sigma_\alpha \sim \text{uniform}(0, 10).$$

The priors for the covariate coefficients in Equation 1 will be

$$\boldsymbol{\beta} \sim \text{normal}\left(\mathbf{0}, 10^2\mathbf{I}\right).$$

The prior density for θ will be informative because previous evidence gives some idea of the size of the effect (Eicher et al., 2005; Gluckman et al., 2005; Shankaran et al., 2005). The parameter θ will reflect neutral, enthusiastic, and skeptical priors (D. Spiegelhalter, Abrams, & Myles, 2004). The prior density will be

$$\theta \sim \text{normal}(\mu_\theta, 0.5^2),$$

where

$$\mu_\theta = \begin{cases} 0.0 & \text{for the neutral prior;} \\ -0.7 & \text{for the enthusiastic prior;} \\ 0.3 & \text{for the skeptical prior.} \end{cases}$$

These values for μ_θ approximate the risk ratios and their 95% credible intervals of the proposal.

6 Posterior Density

The third component is the posterior density, which is determined by the likelihood and the prior. For notational convenience, stack the y_i and s_i into N-vectors $\mathbf{y}$ and $\mathbf{s}$ and the $\mathbf{x}_i'$ into an $N \times p$ matrix $\mathbf{X}$. Also stack the α_s into the S-vector $\boldsymbol{\alpha}$. In an obvious abuse of notation, let normal$(\cdot \mid \mu, \sigma^2)$ denote the normal density function with mean μ and variance σ^2, and uniform$(\cdot \mid a, b)$ denote the uniform density with bounds a and b. Given the observed outcomes $\mathbf{y}$ and associated

sites s and covariates X, Bayes's Theorem yields

$$
\begin{aligned}
&\mathrm{pr}\left(\theta, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mu_\alpha, \sigma_\alpha^2 \mid \mathbf{y}, \mathbf{s}, \mathbf{X}\right) \\
&\quad \propto \prod_{i=1}^{N} \mathrm{pr}\left(y_i \mid \theta, \boldsymbol{\beta}, \alpha_{s_i}, \mathbf{x}_i\right) \mathrm{pr}\left(\theta\right) \mathrm{pr}\left(\boldsymbol{\beta}\right) \mathrm{pr}\left(\alpha_{s_i} \mid \mu_\alpha, \sigma_\alpha^2\right) \mathrm{pr}\left(\mu_\alpha\right) \mathrm{pr}\left(\sigma_\alpha^2\right) \\
&\quad \propto \prod_{i=1}^{N} \exp\left[y_i\left(\alpha_{s_i} + \theta z_i + \mathbf{x}_i'\beta\right)\right]\left[1 + \exp\left(\alpha_{s_i} + \theta z_i + \mathbf{x}_i'\beta\right)\right]^{-1} \\
&\qquad \times \mathrm{normal}\left(\theta \mid \mu_\theta, 0.5^2\right) \times \mathrm{normal}\left(\boldsymbol{\beta} \mid \mathbf{0}, 10^2\mathbf{I}\right) \times \mathrm{normal}\left(\alpha_{s_i} \mid \mu_\alpha, \sigma_\alpha^2\right) \\
&\qquad \times \mathrm{normal}\left(\mu_\alpha \mid 0, 10^2\right) \times \mathrm{uniform}\left(\sigma_\alpha \mid 0, 10\right).
\end{aligned}
$$

The posterior distribution is analytically intractable, but can be calculated by modern Bayesian computational methods. The posterior *marginal* densities of the parameters can be approximated to a high degree of accuracy by Markov chain Monte Carlo (MCMC) methods (Gelman et al., 2004). In essence, MCMC methods produce the marginal densities of each parameter by sampling in a round-robin fashion from the marginal density of each parameter conditional on the values of the other parameters until convergence is achieved. In addition to their posterior densities, all parameters will be summarized by their posterior means, standard deviations, and their 95% credible intervals.

7 Derived Parameters

The posterior density of the adjusted odds ratio comparing treatment to control

$$\omega = \exp(\theta) \tag{3}$$

is obtained by sampling from the posterior density of $\theta \mid \mathbf{y}$.

The posterior density for the proportions of outcomes for either treatment z and any combination of covariates $\mathbf{x}$ and averaged over sites can be obtained as

$$\mathrm{logit}\left[\pi\left(z, \mathbf{x}\right)\right] = \mu_\alpha + \theta z + \boldsymbol{\beta}'\mathbf{x}$$

by sampling from the posterior densities of $\mu_\alpha \mid \mathbf{y}$, $\theta \mid \mathbf{y}$, and $\boldsymbol{\beta} \mid \mathbf{y}$. The posterior densities of the adjusted risk difference

$$\delta(\mathbf{x}) = \pi\left(1, \mathbf{x}\right) - \pi\left(0, \mathbf{x}\right)$$

and the adjusted risk ratio

$$\rho(\mathbf{x}) = \frac{\pi\left(1, \mathbf{x}\right)}{\pi\left(0, \mathbf{x}\right)}$$

for any combination of predictors can then be readily obtained. Again, these derived parameters will be summarized by their posterior means, standard deviations, and their 95% credible intervals.

8 Hypothesis Testing

Adapting the categories provided by D. Spiegelhalter et al. (2004). The primary hypotheses to be tested are whether hypothermia treatment is *beneficial*, *equivalent*, or *harmful* relative to standard care. In addition to the above mutually exclusive categories, hypothermia may be not harmful but nonetheless *nonbeneficial* or not beneficial but nonetheless *nonharmful*. And finally,

hypothermia may be neither beneficial, harmful, nor equivalent, but *equivocal* with respect to standard care.

Let ω be the posterior odds ratio as in Equation 3. Let c_b and c_h with $c_b < c_h$ be clinical thresholds for hypothermia treatment being beneficial or harmful relative to standard care, and let η be a probability level such as .95 or .99. The choice of c_b is based on how large a benefit is required to produce a clinically meaningful result, taking into account the cost and concomitant risks of treatment. The threshold c_h reflects how much harm is allowed before the treatment is deemed clinically inappropriate. The parameter η reflects the uncertainty allowed for each hypothesis. Using odds ratio and these clinical thresholds, hypothermia treatment would be

- beneficial if $\Pr(\omega < c_b) > \eta$,
- equivalent if $\Pr(c_b \leq \omega \leq c_h) > \eta$, or
- harmful if $\Pr(\omega > c_h) > \eta$.

If none of the hypotheses obtain, then hypothermia treatment would be

- nonharmful, i.e., probably equivalent or beneficial, if $\Pr(\omega < c_h) > \eta$, or
- nonbeneficial, i.e., probably equivalent or harmful, if $\Pr(\omega > c_b) > \eta$.

If neither of the hypotheses obtain, then hypothermia treatment would be

- equivocal

with respect to standard care. Multiple choices of clinically relevant criteria can be examined, without requiring statistical adjustments for multiple hypotheses. Similar hypotheses can be tested with $\rho(\mathbf{x})$ and $\delta(\mathbf{x})$.

9 Software

The MCMC analyses will be conducted in WinBUGS (D. J. Spiegelhalter, Thomas, Best, & Lunn, 2002). The WinBUGS program code, provided in the appendix, has already been written and tested. Convergence diagnostics will be conducted by using multiple start values with the potential scale reduction estimator. Artificial data sets resembling the data to be expected have been generated and successfully analyzed. Subsidiary analyses and graphics production will be conducted in R (R Development Core Team, 2006). Connections between R and WinBUGS is provided by the R package `R2WinBUGS` (Sturtz, Ligges, & Gelman, 2005).

10 References

Eicher, D. J., Wagner, C. L., Katikaneni, L. P., Hulsey, T. C., Bass, W. T., Kaufman, D. A., et al. (2005). Moderate hypothermia in neonatal encephalopathy: Efficacy outcomes. *Pediatric Neurology*, *2*(1), 11-17.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, *100*(470), 680–700.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman & Hall.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Gluckman, P. D., Wyatt, J. S., Azzopardi, D., Ballard, R., Edwards, A., Ferriero, D. M., et al. (2005). Selective head cooling with mild systemic hypothermia after neonatal encephalopathy: Multicentre randomised trial. *Lancet*, *365*, 663–670.

R Development Core Team. (2006). *R: A language and environment for statistical computing.* Vienna, Austria. (Available at `http://cran.r-project.org/`)

Shankaran, S., Laptook, A. R., Ehrenkranz, R. A., Tyson, J. E., McDonald, S. A., Donovan, E. F., et al. (2005). Whole-body hypothermia for neonates with hypoxic-ischemic encephalopathy. *New England Journal of Medicine*, *353*(15), 1574-1584.

Spiegelhalter, D., Abrams, K., & Myles, J. P. (2004). *Bayesian approaches to clincal trials and health-care evaluation.* New York, NY: John Wiley & Sons.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2002). *WinBUGS Version 1.4 user's manual.* (Tech. Rep.). Cambridge, UK: Medical Research Council Biostatistics Unit. (Available from `http://www.mrc-bsu.cam.ac.uk/bugs/`)

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*(3), 1–16.

References

1. Badawi N, Kurinczuk JJ, Keogh JM, Alessandri LM, O'Sullivan F, Burton PR, et al. Intrapartum risk factors for newborn encephalopathy: the Western Australian case-control study. BMJ 1998;317(7172):1554-8.
2. Volpe JJ. Neurology of the newborn. 4th ed. Philadelphia: Saunders; 1997.
3. Laptook AR, Corbett RJ. The effects of temperature on hypoxic-ischemic brain injury. Clin Perinatol 2002;29(4):623-49.
4. Gluckman PD, Wyatt JS, Azzopardi D, Ballard R, Edwards AD, Ferriero DM, et al. Selective head cooling with mild systemic hypothermia after neonatal encephalopathy: multicentre randomised trial. Lancet 2005;365(9460):663-70.
5. Shankaran S, Laptook AR, Ehrenkranz RA, Tyson JE, McDonald SA, Donovan EF, et al. Whole-body hypothermia for neonates with hypoxic-ischemic encephalopathy. N Engl J Med 2005;353(15):1574-84.
6. Gunn AJ, Bennet L, Gunning MI, Gluckman PD, Gunn TR. Cerebral hypothermia is not neuroprotective when started after postischemic seizures in fetal sheep. Pediatr Res 1999;46(3):274-80.
7. Gunn AJ, Gunn TR, de Haan HH, Williams CE, Gluckman PD. Dramatic neuronal rescue with prolonged selective head cooling after ischemia in fetal lambs. J Clin Invest 1997;99(2):248-56.
8. Gunn AJ, Gunn TR, Gunning MI, Williams CE, Gluckman PD. Neuroprotection with prolonged head cooling started before postischemic seizures in fetal sheep. Pediatrics 1998;102(5):1098-106.
9. Taylor DL, Mehmet H, Cady EB, Edwards AD. Improved neuroprotection with hypothermia delayed by 6 hours following cerebral hypoxia-ischemia in the 14-day-old rat. Pediatr Res 2002;51(1):13-9.
10. Coimbra C, Wieloch T. Moderate hypothermia mitigates neuronal damage in the rat brain when initiated several hours following transient cerebral ischemia. Acta Neuropathol (Berl) 1994;87(4):325-31.
11. Carroll M, Beek O. Protection against hippocampal CA1 cell loss by post-ischemic hypothermia is dependent on delay of initiation and duration. Metab Brain Dis 1992;7(1):45-50.
12. Colbourne F, Corbett D. Delayed and prolonged post-ischemic hypothermia is neuroprotective in the gerbil. Brain Res 1994;654(2):265-72.
13. Colbourne F, Corbett D. Delayed postischemic hypothermia: a six month survival study using behavioral and histological assessments of neuroprotection. J Neurosci 1995;15(11):7250-60.
14. Szaflarski J, Burtrum D, Silverstein FS. Cerebral hypoxia-ischemia stimulates cytokine gene expression in perinatal rats. Stroke 1995;26(6):1093-100.
15. Hagberg H, Gilland E, Bona E, Hanson LA, Hahin-Zoric M, Blennow M, et al. Enhanced expression of interleukin (IL)-1 and IL-6 messenger RNA and bioactive protein after hypoxia-ischemia in neonatal rats. Pediatr Res 1996;40(4):603-9.

16. Hedtjarn M, Leverin AL, Eriksson K, Blomgren K, Mallard C, Hagberg H. Interleukin-18 involvement in hypoxic-ischemic brain injury. J Neurosci 2002;22(14):5910-9.

17. Zhang RL, Chopp M, Li Y, Zaloga C, Jiang N, Jones ML, et al. Anti-ICAM-1 antibody reduces ischemic cell damage after transient middle cerebral artery occlusion in the rat. Neurology 1994;44(9):1747-51.

18. Zhang RL, Chopp M, Zaloga C, Zhang ZG, Jiang N, Gautam SC, et al. The temporal profiles of ICAM-1 protein and mRNA expression after transient MCA occlusion in the rat. Brain Res 1995;682(1-2):182-8.

19. Plane JM, Liu R, Wang TW, Silverstein FS, Parent JM. Neonatal hypoxic-ischemic injury increases forebrain subventricular zone neurogenesis in the mouse. Neurobiol Dis 2004;16(3):585-95.

20. Busto R, Dietrich WD, Globus MY, Valdes I, Scheinberg P, Ginsberg MD. Small differences in intraischemic brain temperature critically determine the extent of ischemic neuronal injury. J Cereb Blood Flow Metab 1987;7(6):729-38.

21. Minamisawa H, Smith ML, Siesjo BK. The effect of mild hyperthermia and hypothermia on brain damage following 5, 10, and 15 minutes of forebrain ischemia. Ann Neurol 1990;28(1):26-33.

22. Wass CT, Lanier WL, Hofer RE, Scheithauer BW, Andrews AG. Temperature changes of > or = 1 degree C alter functional neurologic outcome and histopathology in a canine model of complete cerebral ischemia. Anesthesiology 1995;83(2):325-35.

23. Kuroiwa T, Bonnekoh P, Hossmann KA. Prevention of postischemic hyperthermia prevents ischemic injury of CA1 neurons in gerbils. J Cereb Blood Flow Metab 1990;10(4):550-6.

24. Baena RC, Busto R, Dietrich WD, Globus MY, Ginsberg MD. Hyperthermia delayed by 24 hours aggravates neuronal damage in rat hippocampus following global ischemia. Neurology 1997;48(3):768-73.

25. Mishima K, Ikeda T, Yoshikawa T, Aoo N, Egashira N, Xia YX, et al. Effects of hypothermia and hyperthermia on attentional and spatial learning deficits following neonatal hypoxia-ischemic insult in rats. Behav Brain Res 2004;151(1-2):209-17.

26. Yager JY, Armstrong EA, Jaharus C, Saucier DM, Wirrell EC. Preventing hyperthermia decreases brain damage following neonatal hypoxic-ischemic seizures. Brain Res 2004;1011(1):48-57.

27. Ginsberg MD, Busto R. Combating hyperthermia in acute stroke: a significant clinical concern. Stroke 1998;29(2):529-34.

28. Grether JK, Nelson KB. Maternal infection and cerebral palsy in infants of normal birth weight. JAMA 1997;278(3):207-11.

29. Impey L, Greenwood C, MacQuillan K, Reynolds M, Sheil O. Fever in labour and neonatal encephalopathy: a prospective cohort study. Bjog 2001;108(6):594-7.

30. Lieberman E, Eichenwald E, Mathur G, Richardson D, Heffner L, Cohen A. Intrapartum fever and unexplained seizures in term infants. Pediatrics 2000;106(5):983-8.

31. Lieberman E, Lang J, Richardson DK, Frigoletto FD, Heffner LJ, Cohen A. Intrapartum maternal fever and neonatal outcome. Pediatrics 2000;105(1 Pt 1):8-13.

32. Petrova A, Demissie K, Rhoads GG, Smulian JC, Marcella S, Ananth CV. Association of maternal fever during labor with neonatal and infant morbidity and mortality. Obstet Gynecol 2001;98(1):20-7.

33. Shalak LF, Laptook AR, Jafri HS, Ramilo O, Perlman JM. Clinical chorioamnionitis, elevated cytokines, and brain injury in term infants. Pediatrics 2002;110(4):673-80.

34. Hajat C, Hajat S, Sharma P. Effects of poststroke pyrexia on stroke outcome: a meta-analysis of studies in patients. Stroke 2000;31(2):410-4.

35. Laptook A, Neonatal Research Network. Adverse outcome increases with elevated temperature for infants provided usual care following hypoxia-ischemic encephalopathy (HIE). Pediatr Res 2006;59:5755.

36. Shalak LF, Laptook AR, Velaphi SC, Perlman JM. Amplitude-integrated electroencephalography coupled with an early neurologic examination enhances prediction of term infants at risk for persistent encephalopathy. Pediatrics 2003;111(2):351-7.

37. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet 2005;365(9467):1348-53.

38. Kuehn BM. Industry, FDA warm to "adaptive" trials. JAMA 2006;296(16):1955-7.

39. Kadane JB. Prime time for Bayes. Control Clin Trials 1995;16(5):313-8.

40. Salmon WC. Rationality and objectivity in science or Tom Kuhn meets Tom Bayes. In: Papineau D, editor. The philosophy of science. Oxford, UK: Oxford University Press; 1996. p. 256-289.

41. Bernardo JM, Smith AFM. Bayesian theory. Chichester, UK: John Wiley & Sons; 1994.

42. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. BMJ 1995;311(7020):1621-5.

43. Hurn PD, Vannucci SJ, Hagberg H. Adult or perinatal brain injury: does sex matter? Stroke 2005;36(2):193-5.

44. Wen TC, Rogido M, Peng H, Genetta T, Moore J, Sola A. Gender differences in long-term beneficial effects of erythropoietin given after neonatal stroke in postnatal day-7 rats. Neuroscience 2006;139(3):803-11.

45. Greenhouse, JB and Wasserman, L. 1995. Robust Bayesian methods for monitoring clinical trials. Stat-Med. 14(12) 1379-91.