STATISTICAL ANALYSIS PLAN

INFLUENCE OF COOLING DURATION ON EFFICACY IN CARDIAC ARREST PATIENTS (ICECAP)

VERSION 1.0 | APRIL 24, 2020

ICECAP Data Coordinating Center
Data Coordination Unit
Department of Public Health Sciences
Medical University of South Carolina
Viswanathan Ramakrishnan, PhD
Ellen Underwood, MS
Sharon Yeatts, PhD

Berry Consultants
Scott Berry, PhD
Kristine Broglio, MS
Joe Marion, PhD

Table of Contents

1 Synopsis of the Study

This statistical analysis plan (SAP) is for the ICECAP trial. ICECAP is a phase II/III adaptive clinical trial, designed to characterize the duration response curve for hypothermia in comatose adult survivors of out of hospital cardiac arrest that have already been rapidly cooled using a definitive temperature control method. Those with and without initial shockable rhythms will be studied as distinct populations.

1.1 Primary Aims

1.1.1 Optimal Duration

To determine, in each of two populations of adult comatose survivors of cardiac arrest (those with initial shockable rhythms and those with PEA/asystole), the shortest duration of cooling that provides the maximum treatment effect as determined by a weighted 90-day modified Rankin score.

1.1.2 Efficacy of Hypothermia

To determine, in each of two populations of adult comatose survivors of cardiac arrest (those with initial shockable rhythms and those with PEA/asystole), whether increasing durations of cooling are associated with better outcomes or recovery implying efficacy of hypothermia compared to no cooling.

1.2 Secondary Aims

1.2.1 Safety

To characterize the overall safety and adverse events associated with duration of cooling

1.2.2 Neuropsycholocal Outcomes

To characterize the effect of duration of cooling on neuropsychological outcomes

1.2.3 Quality of Life

To characterize the effect of duration of cooling on patient reported quality of life

2 Acronyms

Abbreviation	Description
AE	Adverse Event
CCC	Clinical Coordinating Center
CRF	Case Report Form
DCC	Data Coordinating Center
DCU	Data Coordination Unit
DPHS	Department of Public Health Sciences
DSMB	Data and Safety Monitoring board
ITT	Intent to Treat
mRS	modified Rankin Scale
NHLBI	National Heart Lung and Blood Institute
NINDS	National Institute of Neurological Disorders and Stroke
RAR	Response Adaptive Randomization
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation

3 Study Design

This trial is designed as a multicenter, prospective, randomized, phase II/III clinical trial. Comatose adult survivors of out of hospital cardiac arrest that have already been rapidly cooled using a definitive temperature control method (endovascular or surface) will be enrolled in the emergency department or intensive care unit. The maximum number of subjects to be enrolled is 1800.

3.1 Treatment Arms

There are potentially 10 treatment arms defined in the trial:

	Duration of Cooling	Status
1	6	Open if specific conditions are met
2	12	Open at study start
3	18	Open after 'burn-in' period
4	24	Open at study start
5	30	Open after 'burn-in' period
6	36	Open after 'burn-in' period
7	42	Open after 'burn-in' period
8	48	Open at study start
9	60	Open if specific conditions are met
10	72	Open if specific conditions are met

4 Definition of the Target Population and Study Samples

4.1 Target Population

Comatose adult survivors of out of hospital cardiac arrest that have already been rapidly cooled using a definitive temperature control method (endovascular or surface) will be enrolled in the emergency department or intensive care unit. The specific inclusion/exclusion criteria are listed in the protocol.

4.2 Intent to Treat Sample

The intent-to-treat (ITT) sample includes all subjects randomized. All randomized subjects will be analyzed according to the duration to which they were randomized, regardless of the duration actually received.

It is anticipated that subjects may initially be perceived to be of one rhythm type and later found to be of the other rhythm type. In the ITT analyses, these subjects will be included with their corrected rhythm type, but their treatment arm will remain as randomized.

5 General Statistical Considerations

5.1 Subject Accountability

A flowchart will be created to present a summary of participant status.

5.2 Randomization

5.2.1 Procedures

A web-based central randomization system will be developed by the DCC and installed on the WebDCU™ ICECAP study website. Randomization will occur via the study-specific password-protected website accessed by an authorized research coordinator or investigator at the clinical site. If, in rare circumstances, the web system is not available, the coordinator or investigator will have access to emergency randomization procedures that will allow the site to randomize the subject. Upon randomization by the authorized person at each center, an e-mail notification will be sent to the Study EC, Site PI, Site Primary Study Coordinator and relevant CCC and DCC personnel. Subjects will be considered enrolled in this trial at the time of randomization, regardless of whether or not they start or complete study treatment. The specific details concerning randomization will be defined in the ICECAP randomization plan.

5.2.2 Algorithm

The first 200 subjects enrolled overall between the two rhythm types will be randomized equally (1:1:1) between three treatment arms: 12 hours, 24 hours, and 48 hours. After these first 200 patients have been enrolled, patients will be allocated to treatment arms (durations 12 through 48) based on response adaptive randomization (RAR). RAR will be applied separately within each rhythm type. Interim analyses will be conducted approximately every 50 enrollments in order to update the response adaptive randomization probabilities. The response adaptive randomization probabilities determined at each interim analysis will be applied for the newly enrolling subjects until the next interim analysis and update, one month later.

The 6, 60, and 72-hour duration treatment arms are initially closed to enrollment, but can be opened for randomization based on the accruing evidence of efficacy. The 6-hour duration treatment arm will be opened for a rhythm type if there are more than 100 subjects enrolled across all arms in that rhythm type and there is at least a 0.33 probability that 6 hours is the

target duration for that rhythm type. The 60 and 72-hour duration treatment arms will be opened incrementally. These arms open on either rhythm if there is at least a 0.33 probability that the target duration for that rhythm type is at or above that next shorter duration.

5.3 Blinding

Subjects themselves will be comatose during the intervention period. It is not practicable to blind the clinical care team or the subject's family to the duration of cooling. The primary outcome assessment in this trial will be performed by a study team member blinded to treatment.

5.4 Missing Data

The primary analysis uses the ITT sample which includes all randomized subjects. Extensive effort should be put forth to ensure near complete follow-up, especially with regard to the primary outcome and occurrence of death. To some extent, however, missing data is inevitable, due to decline/withdrawal of consent and/or Loss-to-Follow-Up.

At each interim analysis, and at the final analysis, there will be subjects who have an unknown 90-day mRS value. We use the 30-day mRS value in a longitudinal model, as possibly predictive of the 90-day mRS, allowing subjects with this earlier measurement to be included in the analyses of the 90-day measurement. This longitudinal model allows for the accruing empirical data to determine the strength of the association between the two values (30-day and 90-day) for each treatment arm and rhythm type. Analyses of the 90-day mRS values are performed with multiple imputation from the longitudinal model for patients with an unknown 90-day mRS value. This is done for the purpose of fitting the models with maximum information at every interim stage, to determine the probabilities of following dose assignments in the adaptive design. Details of this model are provided in Section 6.2.2.

5.5 Comparability of Patients Randomized across Duration Arms

A description of the baseline characteristics of trial participants will be presented by duration group. Dichotomous variables will be summarized as number (%). Percentages will be calculated based on the number of participants with available data for that variable. Continuous variables will be summarized by the mean and standard deviation (SD). In the case of variables with missing values, the denominator will be stated in the summary table or in a footnote to the summary table.

5.6 Interim Analysis Process for Design, Model Checking, and Dissemination of Information

Prior to executing the interim analysis process (including the response adaptive update) at each interim analysis, the following procedure will be used:

1. Data will be cleaned to the greatest reasonable level, and the database frozen. Specifically the key data elements are the rhythm type, treatment assignment level; mRS at 30 days and 90 days.

2. The data will be transferred to the unblinded statistical team for analysis. The interim analysis will be conducted by the unblinded statistical team, using the executable file developed and validated by Berry Consultants.

3. The resulting randomization vector will be entered in WebDCU by the unblinded statistical team. Once submitted, the updated allocation probabilities will be used for subsequent randomizations.

4. An abridged DSMB Interim Analysis report, including the results of the primary analysis and an indication as to whether a stopping rule has been met, will be assembled and submitted for review by the DSMB Chair and Statisticians. If the analysis indicates that one of the stopping rules has been metthe unblinded statistical team will prepare a full DSMB report to submit to the DSMB for review and evaluation.

6 Primary Efficacy Analysis

6.1 Overview

Provided here is a summary of how the outcome variable will be analyzed. The design is a Phase II/III adaptive clinical trial. The purpose of the trial is to determine the shortest duration of cooling that provides the maximum treatment effect, as determined by a weighted 90-day mRS and whether increasing durations of cooling are associed with better outcomes or recovery. Response adaptive randomization will be used to favor the duration most likely to be the target and arms with greater variability around the primary endpoint or a smaller sample size. The primary analysis will use the ITT sample (Section 4.2) to evaluate the model relating the primary endpoint to cooling duration and is estimated for each rhythm type.

6.2 Primary Outcome Variable(s) Analysis

6.2.1 Definition of Outcome

The primary endpoint is the 90-day mRS (M_{90}). The primary analysis uses a weighting of the values as shown.

$$W(M_{90}) = \begin{cases} 10 & M_{90} = 0 \\ 9 & M_{90} = 1 \\ 8 & M_{90} = 2 \\ 6 & M_{90} = 3 \\ 0 & M_{90} = 4,5,6 \end{cases}$$

Let the observed weighted 90-day mRS for subject *i* be W_i.

6.2.2 Handling of Missing Outcome Data

At each interim analysis, and at the final analysis, there will be subjects who have an unknown 90-day mRS value. We use the 30-day mRS value as possibly predictive of the 90-day mRS. This longitudinal approach allows for the accruing empirical data to be incorporated in the calculation of the posterior probabilities as determined by the strength of the association between the two values (30-day and 90-day) for each treatment arm and rhythm type. Analyses of the 90-day mRS values are performed with Bayesian multiple imputation from the longitudinal model for patients with an unknown 90-day mRS value.

The longitudinal approach proposed will map the 7 possible 30-day mRS values to the 7 possible 90-day mRS values, assuming a Markovian structure for the 'transitions' from the 30-day mRS state (k) to the 90-day mRS state (j). The transition matrix, P, represents the matrix of probabilities for a subject transitioning from the 30-day mRS state (rows) to the 90-day mRS state (columns). For example, p_{32} is the probability that a subject that is a 30-day mRS of 3 becomes a 2 at 90 days.

$$P = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{05} & p_{06} \\ p_{10} & p_{11} & & p_{15} & p_{16} \\ & & \ddots & & \\ \vdots & \vdots & p_{kj} & \vdots & \vdots \\ & & & \ddots & \\ & & & & \\ p_{60} & p_{61} & & p_{65} & p_{66} \end{bmatrix}$$

A Dirichlet prior distribution is placed on each vector of probabilities from each state at 30 days to each state at 90 days.

$$(p_{k0} \quad \cdots \quad p_{k6})^T \sim Dirichlet((\alpha_{k0} \quad \cdots \quad \alpha_{k6})^T) \text{ for k =0, ..., 6}$$

An assumption is that all transitions from a 30-day mRS of 6 are to a 90-day mRS of 6 (death at 30 days is death at 60 days). For all treatment arms and both rhythm types, the following prior weights for the transitions are assumed:

$$\alpha_{kj} = \begin{cases} 0.1, & k \neq 6 \\ 0, & k = 6, j \neq 6 \\ 1, & k = 6, j = 6 \end{cases}$$

Each prior weight can be interpreted as the number of subjects that have made that particular transition. These prior weights will be quickly overwhelmed by the accruing data and yet provide reasonable estimates if the data are minimal for an mRS state.

The probability vectors have separate posterior distributions by treatment arm and rhythm type. The observed transitions for the same treatment arm *h* and rhythm type *r* contribute fully to that particular posterior distribution, while the transitions from other treatment arms and for other rhythm types contribute 1/4 of their full weight to the posterior distribution. Thus, there is borrowing of partial information from other treatment arms and the opposite rhythm type.

Let the number of subjects in the trial that transition from mRS state *k* at 30 days to mRS state *j* at 90 days, for treatment arm *h* and rhythm type *r*, be $T_{h,r}(k,j)$. The posterior distribution for the probability vector from mRS state *k*, for treatment arm *h* and rhythm type *r*, is modeled as

$$Z_{h,r} = \begin{pmatrix} P_{k0} \\ \vdots \\ P_{kj} \\ \vdots \\ P_{k6} \end{pmatrix} = Dir \begin{pmatrix} \alpha_{k0} + T_{h,r}(k,0) + \frac{1}{4} \sum_{h',r' \in (h,r)'} T_{h',r'}(k,0) \\ \vdots \\ \alpha_{kj} + T_{h,r}(k,j) + \frac{1}{4} \sum_{h',r' \in (h,r)'} T_{h',r'}(k,j) \\ \vdots \\ \alpha_{k6} + T_{h,r}(k,6) + \frac{1}{4} \sum_{h',r' \in (h,r)'} T_{h',r'}(k,6) \end{pmatrix}$$

where (*h*,*r*)' refers to all treatment arms and the rhythm type that are different than the (*h*,*r*) pair.

6.2.3 Primary Model of Outcome

The mean weighted 90-day mRS value for a given duration *h*, $\bar{W}_{h,r}$, is modeled as normally distributed

$$\left[\bar{W}_{h,r}\right] \sim N\left(\theta_{h,r}, \frac{\sigma_r^2}{n_{h,r}}\right)$$

where the treatment arm is *h* and the rhythm type is *r*. The variance components are modeled separately for each rhythm type, with weak prior distributions

$$[\sigma_r^2] \sim InverseGamma(2.5, 22.5).$$

The mean weighted 90-day mRS is modeled via an inverted-U shape duration response model. For each rhythm type, a separate and identical instance of the model is used.

$$\theta_h = \begin{cases} \beta_0 + \beta_1 h^{\beta_3} & h \le \gamma_1 \\ \beta_0 + \beta_1 \gamma_1^{\beta_3} & \gamma_1 < h \le \gamma_2 \\ \beta_0 + \beta_1 \gamma_1^{\beta_3} - \beta_2 \left(h - \gamma_2\right)^{\beta_4} & \gamma_2 < h \end{cases}$$

The values of θ_h will be restricted to be between 0 and 10. The parameters γ_1 and γ_2 are referred to as the change-points. The parameter γ_1 represents the change point between the increasing phase and the plateau phase. The duration response curve is "flat" between γ_1 and the second change-point, γ_2, and the duration response curve decreases after γ_2. An important aspect of the model is that the change-points can be smaller than the minimum cooling duration, h=1 (6 hours), or greater than the maximum cooling duration, h=10 (72-hours), thus allowing the curve to be increasing, decreasing, or flat over the entire range of cooling durations. The model has the following constraints: $\gamma_1 < \gamma_2$, $\beta_1, \beta_2, \beta_3, \beta_4 > 0$.

Because the modeling is done separately for each rhythm type, a superscript *r* is added on the parameters to notate the specific rhythm type. The prior distributions for the change-points are defined on the space where $\gamma_1^r < \gamma_2^r$ as follows:

$$\left[\gamma_1^r\right] \propto \mathrm{N}(4, 10^2) \text{ for } r = 1, 2$$

and

$$[\gamma_2^r] \propto N(8, 3^2) \text{ for } r = 1,2.$$

The following priors for each instance are utilized:

$$[\beta_0^r] \sim N(4, 4^2) \text{ for } r = 1,2$$
$$\left[\beta_j^r\right] \propto 1 \text{ for } j = 1,\ldots,4; r = 1,2$$

6.2.4 Quantities of Interest

6.2.4.1 Posterior Probability of Being the Target Duration

The γ_1 parameter is interpreted as the theoretical optimal duration of cooling, the shortest duration that achieves the maximum treatment effect. The *target duration* is defined based on

γ_1 and γ_2. The target duration is the shortest duration greater than γ_1, if γ_1 is less than 72 hours, or the longest duration if γ_1 is greater than 72 hours.

The posterior probability that each treatment arm is the target duration is calculated for each rhythm type. For each treatment arm *h* and rhythm type *r*, let $\Pr_{h,r}$ be the probability that treatment arm *h* is the target duration for rhythm type *r*. The treatment arm with the highest posterior probability of being the target duration for a particular rhythm type is the most likely target duration for that rhythm type; the corresponding duration and the posterior probability are denoted by *h** and $\Pr_{h,r}^{h^*}$.

6.2.4.2 Posterior Probability Superior to Smaller Duration

For each treatment arm h and rhythm type r, we calculate $E\left(\theta_{h,r}\right)$, the posterior mean weighted 90-day mRS and $V\left(\theta_{h,r}\right)$, the posterior variance of $\theta_{h,r}$. Within each rhythm type *r*, and for each duration *h*, we calculate the posterior probability that it has a superior mean weighted 90 day mRS as compared to each treatment arm with a shorter cooling duration. Because of the form of the duration-response model, it is sufficient to calculate only the posterior probability that a treatment arm is superior to the 6-hour duration arm. We refer to the posterior probability that each treatment arm is superior to a shorter duration treatment arm as $\Pr\left(\theta_{h,r} > \theta_{1,r} \mid data\right)$.

At the final analysis at the end of the trial, for the determination of a positive treatment effect, the posterior distribution using a hierarchical model is used. This hierarchical model allows for the 'borrowing' of information across the rhythm types to inform the determination of efficacy. The following hierarchical prior structure is used, with the restriction that $\gamma_1^r < \gamma_2^r$:

$$\left[\gamma_1^r\right] \propto \mathrm{N}(\mu_1, \tau^2) \text{ for } r = 1, 2$$

and

$$\left[\gamma_2^r\right] \propto \mathrm{N}(\mu_2, \tau^2) \text{ for } r = 1, 2$$

With hyperparameters

$$[\mu_1] \sim N(4, 10^2)$$
$$[\mu_2] \sim N(8, 3^2)$$
$$[\tau^2] \sim G^{-1}(0.05, 0.001).$$

The conclusion that cooling duration h^* is effective in rhythm type r is made if the posterior probability that the mean weighted 90-day mRS for arm h^* is greater than the mean weighted 90-day mRS for a duration shorter than h^* is greater than 0.975. That is, for some $h<h^*$,

$$\Pr\left(\theta_{h^*,r} > \theta_{h,r} \mid data\right) > 0.975 .$$

6.2.5 Model Calculations

Each of the quantities of interest are calculated using a Bayesian Markov chain Monte Carlo algorithm. The algorithm successively samples observations for the posterior distribution. A single chain of length 100,000 is used, with a burn-in sample of 10,000. The algorithm has the following structure:

1. Initial values are used for each parameter.
2. Observations from each complete conditional distribution are drawn, one at a time, successively. For many of these a Metropolis-Hastings algorithm is used for this draw.
3. After the burn-in:
    a. Each sampled value of $\theta_{h,r}$ are stored to calculate the mean and variance.
    b. From each draw from the posterior the target dose is recorded for each rhythm type (it is a single unique dose for each posterior sample). The frequency of draws in which each duration is the target is the estimate of the posterior probability of each dose being the target; $\Pr_{h,r}$
4. From each draw from the posterior it is recorded for each dose if it is superior to the 6-hour duration. The frequency of the 100,000 draws in which a duration is superior to the 6-hour dose is the estimate of the posterior probability of superiority to the 6-hour duration; $\Pr\left(\theta_{h,r} > \theta_{1,r} \mid data\right)$

6.2.6 Analyses

The first interim analysis will occur after the first 200 subjects have been enrolled. Subsequent interim analyses will be conducted approximately every 50 enrollments, or approximately monthly. The implementation of these randomization updates does not require DSMB approval; however the DSMB will be notified of each interim analysis and the corresponding update.

6.2.7 Criteria for Stopping Accrual

6.2.7.1 Stopping for Expected Futility

At each interim analysis, the trial may stop for futility if no cooling duration greater than 6 hours is found to be more effective than the 6-hour duration. Futility will be assessed separately for

each rhythm type. Therefore, the trial could be declared futile for one rhythm type, and yet continue to enroll subjects of the opposite rhythm type. If both rhythm types are not stupped for futility, the trial will continue to enroll to the maximum sample size. Specifically, a rhythm type will stop for futility if

1. At least 50 patients have been randomized to the 6-hour duration arm
2. There is at least a 50% probability that the 6-hour duration treatment arm is the target duration

6.2.7.2 Stopping for Expected Success

The trial will not be stopped early for success.

6.2.8 Response Adaptive Randomization (RAR)

The first 200 subjects enrolled overall between the two rhythm types will be randomized (1:1:1) between three treatment arms: 12 hours, 24 hours, and 48 hours. After these first 200 patients have been enrolled, additional arms between 12 and 48 hours (18, 24, 30, 36, and 42) will be opened, and subjects will be allocated based on response adaptive randomization (RAR), conducted separately within each rhythm type. Once RAR begins, interim analyses will be conducted every 50 subjects in order to update the randomization probabilities, which will then be applied to subsequently enrolled subjects until the next interim analysis and update. Shorter and longer durations may also be opened. If the emerging duration-response curve is flat, then a 6-hour arm will be opened. For a given rhythm type, the 6-hour duration arm will be opened if there are more than 100 subjects enrolled across all arms and there is at least a 0.33 probability that 6 hours is the target duration. If the data suggest increasing efficacy with increasing duration through 48 hours, then the 60 and 72-hour duration arms will be opened incrementally. These arms open on either rhythm if there is at least a 0.33 probability that the target duration for that rhythm type is at or above that next shorter duration.

Randomization probabilities are weighted according to the posterior probability that each treatment arm is the target duration. The goal of the adaptive randomization is to allocate subjects to the arms most likely to be the target duration but also to learn effectively about the duration response curve.

The randomization probabilities for each treatment arm within each rhythm type are proportional to $\Pr_{h,r}$, the probability that a treatment arm is the most likely target duration for a rhythm type *r*.

$$q_{h,r} = \frac{\Pr_{h,r}\,\delta_{h,r}}{\sum_{l=1}^{10}\Pr_{l,r}\,\delta_{l,r}} \text{ for } h = 1,2,\ldots,10$$

The treatment arms 6-hours (*h*=1), 60 hours (*h*=9), and 72 hours (*h*=10) will initially be closed, but may be later opened according to the randomization rules above. If a treatment arm is open we set $\delta_{h,r}$ = 1, and if a treatment arm is closed we set $\delta_{h,r}$ = 0.

7 Subgroup Analysis of the Primary Efficacy Outcome

At the end of the trial, analyses for important subgroups (gender, age strata, pre-existing comorbidities including diabetes, malignancy, prior neurological disease) will be conducted within each rhythm stratum for the primary endpoint.

8 Safety Analyses

8.1 Overview

The Medical Safety Monitor (MSM) and Data and Safety Monitoring Board (DSMB) will receive periodic safety reports of all adverse events including serious adverse events (SAEs). All AEs and SAEs will be summarized by preferred term and associated system-organ class according to the MedDRA adverse reaction dictionary and by treatment group in terms of frequency of the event, number of subjects having the event, severity, and relatedness to the study treatment.

8.2 Definition of Safety Outcomes

This patient population presents with significant morbidity; as such it is important to evaluate the presence of events with respect to temporal relationship to treatment (i.e. novel onset or worsening) as well as its relationship across durations. The events defined below are likely and anticipated.

Adverse Event	Expected Incidence
Pneumonia	40-60%
Other infections (including UTI and bacteremia)	17-40%
Malignant cardiac arrhythmia	46-65%
Seizures	20-40%
Neurological worsening	30-50%
Electrolyte abnormalities	70-80%
Venous thrombotic disease	10-20%

Coagulopathies	10-20%

All AEs and SAEs are summarized by preferred term and associated system-organ class according to the MedDRA adverse reaction dictionary and by treatment group in terms of frequency of the event, number of subjects having the event, time relative to randomization, severity, and relatedness to the treatment. Cumulative incidences of the specific SAEs related to treatment, as well as all SAEs, will be compared across arms.

8.3 Monitoring Methods

8.3.1 Un-Blinded Statistical Monitoring

The events of interest listed in the table above will be specifically monitored in the DSMB reports. The pooled overall rate, as well as the rate by duration within each rhythm type, will be presented.

8.4 Analysis of 90-Day All-Cause Mortality

Models will include duration and rhythm type as the primary independent variables; the corresponding interaction will also be considered. Site will be included as a random effect.

8.4.1 Endpoint

Mortality will be identified via reason for end-of-study. The time from randomization to either event or censoring occurrence will be calculated. Subjects who do not experience the event during the follow-up period will be censored at the time of their last follow-up or at 90 days, whichever occurs first. Subjects who withdraw consent will be censored at the date of consent withdrawal. Subjects who are lost to follow-up will be censored as of their last follow-up assessment.

8.4.2 Analysis

Mortality will be analyzed via survival analysis using a Cox proportional hazards model. We will also perform an analysis to compare the selected duration to the lowest duration, after adjusting for other relevant prognostic variables and demographics.

8.5 Analysis of mRS ≥4 at 90 Days

8.5.1 Endpoint

The 90-day mRS will be dichotomized (mRS ≥4 vs otherwise).

8.5.2 Analysis

This binary outcome will be analyzed using a generalized linear model. We will also perform an analysis to compare the selected duration to the lowest duration, after adjusting for other relevant prognostic variables and demographics.

9 Secondary Efficacy Analyses

9.1 Data Analysis Model and Specification of Secondary Outcomes

This study is designed to achieve the primary aims. However, it also offers the opportunity to conduct analysis to address important secondary objectives, including the effect of treatment on mortality, patient-reported, and neuroposychological outcomes. The primary manuscript will include reporting of the composite outcomes derived from the NIH Toolbox described in Section 9.3 below. Analyses for important subgroups (gender, age strata, pre-existing comorbidities including diabetes, malignancy, prior neurological disease) will be conducted within each rhythm stratum for these composite endpoints.

Analyses of the individual neyropsychological and patient-reported outcomes will be reserved for a secondary manuscript focused on these outcomes; this is specified not to diminish their importance but to allow a full-depth analysis and discussion which would not be possible as a component of the primary manuscript.

It is hypothesized that, among subjects who wake, duration of cooling may have an impact on cognitive function and/or quality of life. If the effect of cooling is such that it decreases mortality but leaves those patients with at least moderate disability, it may appear that cooling worsens cognition and/or quality of life. Therefore, the interpretation of these results must consider the effect of cooling on the availability of subjects for cognitive testing. Subjects with missing data will be excluded.

Analyses of secondary outcomes are considered supportive in nature; as such, the focus is on point estimates and confidence intervals rather than p-values. Unless otherwise specified, no multiplicity adjustment for these outcomes is planned.

All models will include duration and rhythm type as the primary independent variables; the corresponding interaction will also be considered. Site will be included as a random effect.

9.2 Effect of Cooling Duration on Patient-Reported Outcomes

Endpoints will be analyzed as continuous dependent variables using generalized linear models. We will also perform an analysis to compare the selected duration to the lowest duration, after adjusting for other relevant prognostic variables and demographics.

9.3 Effect of Cooling Duration on Neuropsychological Outcomes

The NIH Toolbox tests can be subdivided into crystallized (i.e., general knowledge base) and fluid (i.e., thinking and reasoning) measures, providing information about both patients' premorbid and current functioning. A fluid composite score will be obtained for fluid measures (i.e., those expected to change with injury). A stability composite score will be calculated for crystallized measures (i.e., those not expected to change with injury).

Endpoints will be analyzed as continuous dependent variables using generalized linear models. We will also perform an analysis to compare the selected duration to the lowest duration, after adjusting for other relevant prognostic variables and demographics.

10 Sample Size Determination

This trial will enroll a maximum of 1800 patients. If the trial is not stopped early for futility, it will continue to enroll to the maximum sample size. Extensive numerical simulations of the design were conducted over a range of potential scenarios, based on which, a maximum of 1800 patients was deemed sufficient for the primary analysis. See the simulation report included as an appendix to the protocol for full simulation results, including sample sizes ranging from 1500 to 2300 under a variety of reference scenarios.

Simulation Scenarios used to Study Operating Characteristics: To demonstrate the variety of duration–response scenarios considered in the simulation study, nine examples are provided in Figure 1. These include strong treatment effect (5a), U-shaped effect (5b), Increasing linearly (5c), Large effect at a longer duration (5d), More than six hours is harmful (5e), Early plateau (5f), Similar dose response for the two rhythms (red and black) early effect (5g) and late effect (5h), and contrasting dose responses for the two rhythms (5i). For each simulated scenario, up to three durations were determined to be clinically accurate selections: the shortest duration that achieves the maximum treatment effect and up to two more durations that are clinically similar (i.e., within 12 hours of the optimal duration and achieving at least 70% of the maximum treatment effect).

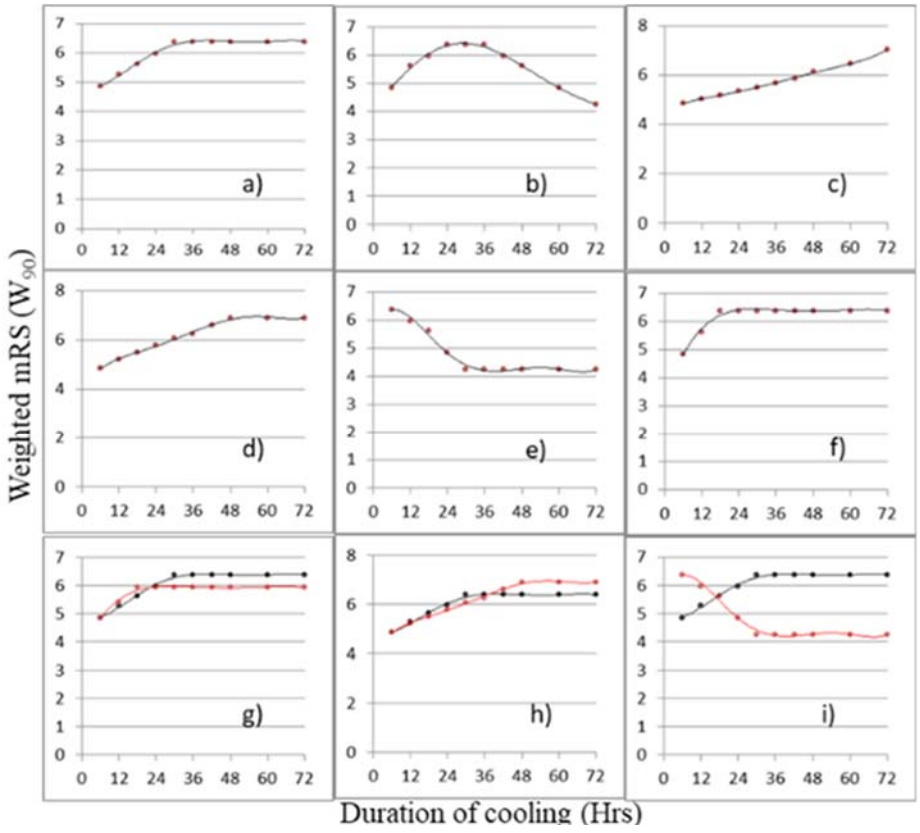

Figure 1. Duration-response Scenarios

Under the Bayes framework, we considered the posterior probability of "success" using two criteria: i) ability to select an acceptable target (SAT) duration larger than the lowest duration enrolled (initially 12 hrs and possibly 6hrs) and ii) there is a positive response (Pos) at a given duration. Although, ideally one would desire (which is achievable), we believe that achieving either one of them would amount to a positive trial.

For each simulated scenario, whether the treatment effect for any duration is greater than for a shorter duration was tested. In certain situations, the design may have convincing evidence of response, but may not be able to definitively choose a duration (e.g. a gently upsloping with plateau, as in Figure 5d). Conversely, the design may be able to choose a target duration, but may not be able to definitively demonstrate duration response (example: true target duration 12 hours, but trial results are insufficient to declare 12 hours is superior to 6 hours). Here, the objective is to calculate the probability of concluding that there is a positive duration response

curve, regardless of the target duration and whether it is correctly selected, in any simulation scenario which specifies an increase in treatment effect with increasing duration. By assigning common priors, the Bayesian statistical model building will be able to utilize information from both rhythm types in the determination of the characteristics of the duration response curve. This increases efficiency of each model fit and therefore leads to a smaller sample size than the approach that would treat the two rhythm types independently. In the Bayes framework, posterior probabilities larger than 70% for SAT or Pos would be considered high.

Sample Size Justification: The proposed sample size was chosen by evaluating the effect of varying sample size from 1,500 to 2,300 on the probability of identifying a positive duration response and the probability of detecting the best duration in simulations of 3 reference scenarios. Details are found in Section 4.7 of the enclosed simulation report and are shown graphically in Figure 2. The reference scenarios consider a 1.5 difference in mean weighted mRS (approximately 16%) treatment effect for both shockable and non-shockable rhythms based on conservative interpretation of the only two RCT comparing cooling to no cooling (these trials used 12 and 24 hour durations of cooling respectively to achieve absolute increases of 16-23% in the proportion of patients with a good neurological outcome after cardiac arrest with an initial shockable rhythm). The treatment effect in the 3 reference scenarios plateaus at 18, 30, or 48 hours. The sample size of 1,800 was selected based on consideration of all reference scenarios, and two additional reference scenarios with weak treatment effects. In the middle reference scenario with a plateau at 30 hours, a sample size of 1800 provides a probability of 0.94 for identifying a positive duration response and 0.78 for correctly identifying the shortest effective duration. The duration response curve associated with this scenario, and the underlying mRS distribution (including mortality) is demonstrated in Figure 3. In the most challenging, early (18 hour) plateau reference scenario, a sample size of 1800 is the minimum required to provide a

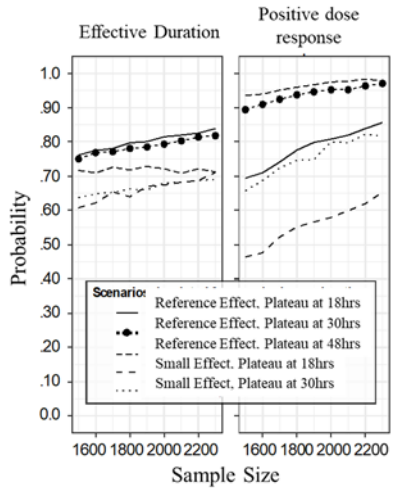

Figure 2. Sample size justification

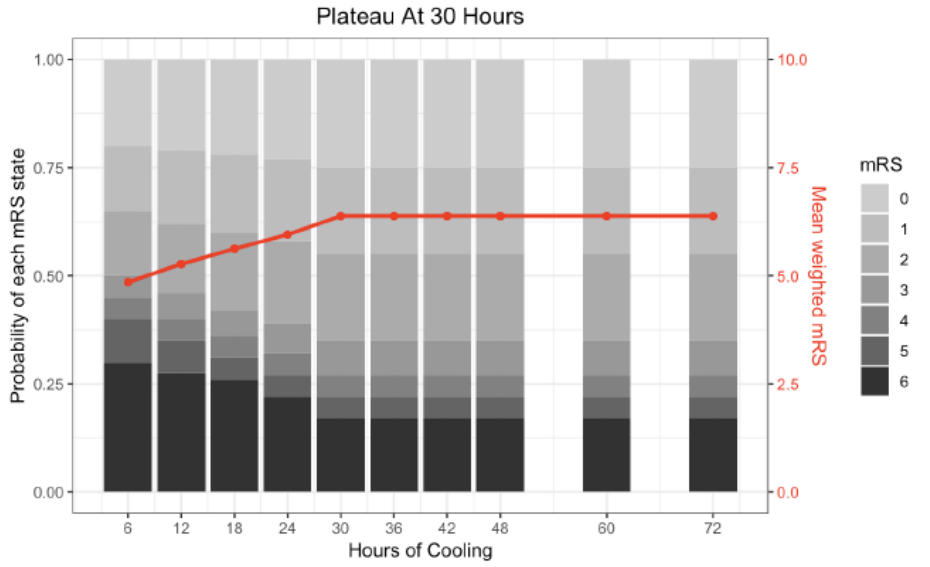

Figure 3. Neurological outcome, including mortality, by duration of cooling

probability of 0.78 for identifying a positive duration response and 0.80 for correctly identifying the shortest effective duration. To prevent extreme imbalances that would render either rhythm type noninformative, we have capped enrollment of either type to 70% of the maximal sample size.

Scenario	Shockable Rhythm			Non-Shockable Rhythm		
	SAT	Pos	Fut	SAT	Pos	Fut
a) Strong Effect (30hrs)	0.78	0.94	0.00	0.78	0.94	0.00
b) U-shaped response	0.83	0.70	0.01	0.82	0.70	0.01
c) Increasing response	0.39	0.85	0.01	0.40	0.85	0.02
d) Late Plateau (48 hrs)	0.78	0.99	0.00	0.76	0.99	0.00
e) Over 6hrs not good	0.93	0.00	0.96	0.92	0.00	0.96
f) Early Plateau (18hrs)	0.80	0.78	0.00	0.80	0.77	0.00
g) Same Rhythms-early	0.79	0.84	0.02	0.65	0.73	0.02
h) Same Rhythms-late	0.77	0.96	0.00	0.76	0.98	0.00
i) Opposing Rhythms	0.84	0.53	0.00	0.91	0.00	0.92

SAT = Pr(Select Acceptable Target)
Pos = Pr(Positive Response)
Fut = Pr(Futility)

Bayes posteriors Total n=1800

Sensitivity of Operating Characteristics to Alternative Duration Response Scenarios at N=1800: With a maximum of 1800 patients, assuming 50% are in each rhythm type, the following Table provides the posterior probabilities for i) and ii) mentioned above, along with the probability of futility for the scenarios presented in Figure 3. With this sample size, as shown in Table 2, for all relevant scenarios, the posterior probabilities of SAT or Pos are larger than 70%. In most of the scenarios, both probabilities are high.

The full design report is included as an appendix to the ICECAP protocol.