Intracerebral Hemorrhage Deferoxamine Trial (iDEF Trial)

STATISTICAL ANALYSIS PLAN

Prepared by

Sharon D. Yeatts, Ph.D.
Yuko Y. Palesch, Ph.D.
Lydia D. Foster, MS

Data Coordination Unit

Department of Public Health Sciences
Medical University of South Carolina
Charleston, SC 29425

Version 1.1

TABLE OF CONTENTS

1. LIST OF ABBREVIATIONS

AE	Adverse Event
AEOSI	Adverse Events of Special Interest
ARDS	Acute Respiratory Distress Syndrome
DCR	Data Clarification Request
DFO	Deferoxamine Mesylate
SDMC	Statistical and Data Management Center
DSMB	Data and Safety Monitoring Board
GCP	Good Clinical Practice
HR	Hazard Ratio
ICH	Intracerebral Hemorrhage
ITT	Intent-to-Treat
MoCA	Montreal Cognitive Assessment
NIHSS	National Institute of Health Stroke Scale
SAE	Serious Adverse Event
SIS-16	Stroke Impact Scale-16

2. STATISTICAL ANALYSIS PLAN AND STATISTICAL REPORTS

This document provides the details of statistical analyses planned for the iDEF Trial. In addition, it discusses the statistical issues relevant to these analyses (e.g., sample data to be used, missing data, adjustments for multiplicity, etc.)

The partially unblinded statistician at the Statistics and Data Management Center (SDMC) generates closed and open DSMB Reports semiannually. Each semiannual report provides cumulative summary statistics on enrollment; subject status in the study (e.g., number completed 90 day assessments); baseline characteristics; protocol violations; safety data, including AEs and SAEs by AE code, severity, and relatedness to the study medication; and data management/quality information (e.g., timeliness and completeness of data entry by the clinical centers via the iDEF Trial Website; number of DCRs generated and resolved). The statistics for the closed DSMB Reports are provided by partially unblinded treatment group (A/B). The open report contains aggregate statistics only, i.e., not by treatment group.

3. OBJECTIVES OF THE STUDY

3.1. Efficacy

The primary objective of the iDEF Trial is to assess the futility of deferoxamine mesylate (DFO), when initiated within 24 hours of symptom onset and administered for 3 days, versus placebo in subjects with ICH. The primary futility hypothesis is that DFO will increase the proportion of subjects with a good functional outcome, as defined by the modified Rankin Scale (mRS) 0-2 at 90 days, by less than absolute 12% when compared to placebo.

In addition, the Trial plans to evaluate the effect of DFO as compared to placebo based on the proportion of subjects with mRS 0-3. For supportive evidence, other clinical outcome measures at 3 months, including the National Institutes of Health Stroke Scale (NIHSS), Montreal Cognitive Assessment (MoCA), and the Stroke Impact Scale-16 (SIS-16), will also be assessed.

3.2. Safety

The safety of DFO when compared to placebo is evaluated by mortality, SAEs, and Adverse Events of Special Interest (anaphylaxis, unexplained hypotension requiring medical intervention, development of new and unexplained visual or auditory changes, and ARDS).

4. STUDY DESIGN

The study is of a two-arm parallel futility design whereby eligible subjects are randomized in a 1:1 ratio to either the DFO group or to the placebo group. Each subject is followed for 180 days from randomization.

5. SAMPLE SIZE DETERMINATION FOR THE PRIMARY EFFICACY ANALYSIS

The total sample size for the futility threshold of 12% (the absolute increase of DFO over the placebo arm in the proportion of subjects with good outcome), assuming a placebo proportion of 28%, and Type I and Type II error probabilities of 0.10 and 0.80, is 254 subjects.

The primary analysis will be conducted according to a modified intent-to-treat (ITT) principle, wherein subjects in whom the study drug infusion is not initiated will be excluded from the analysis. Therefore, the final sample size was inflated by a factor of 1.11 (Friedman, Furberg, and DeMets) to account for dilution of the treatment effect associated with a conservative drop-out rate of 5% (due to loss-to-follow-up (LTFU) and withdrawal of consent), as well as an anticipated <4% of randomized subjects in whom the study drug is not initiated. Therefore, the final sample size is 294 subjects.

6. DEFINITION OF TARGET POPULATION AND STUDY SAMPLES

6.1. Target Population

The target population to which the DFO treatment regimen may be applied are patients with ICH in whom study drug can be initiated within 24 hours of symptom onset.

6.2. Modified Intent-to-Treat Sample

As the primary analysis, all efficacy and safety outcome measures are analyzed under the modified ITT principle. Under this principle, the evaluable sample includes all randomized subjects in whom the study infusion is initiated, regardless of whether or not it was prematurely discontinued. Missing outcome data for subjects included in the modified ITT sample are handled as specified in Section 10.

7. RANDOMIZATION

The randomization takes place centrally via the iDEF Trial Website via a combination of minimization and biased coin methodologies. Subjects are randomized in a 1:1 ratio (DFO: placebo), controlling for ICH score, onset-to-treatment (OTT) time, ICH volume, NIHSS score, concurrent use of warfarin at the time of ICH onset and clinical site. The computer program

developed at the SDMC makes the treatment assignment based on the current status of treatment group distribution and study drug availability. The detailed randomization scheme and source codes are provided in the Randomization Plan document.

The randomization is never deterministic. Although a randomization scheme with any constraints would yield some bias in the inferences from using standard analytic methods, Efron (1971) shows the appropriateness of standard statistical tests under the biased coin randomization in large studies. Friedman et al (1998 p. 72) note that the variance terms under the biased coin design tend to be larger than under simple randomization. The consequence is that it would be more difficult to reject the null hypothesis, and therefore, we would be more conservative in determining the futility of the treatment.

A "Real-Time" randomization procedure is implemented via the iDEF Trial Website on the WebDCU™ System, where the clinical center staff enters the basic baseline (e.g., ICH score, anticipated time to treatment, etc) and eligibility information of a subject prior to enrollment. If the subject's eligibility status is confirmed, the computer program on the WebDCU™ server evaluates the treatment arm distribution and selects a randomization number based on the randomization scheme. The randomization number corresponds to a specific label in the Randomization Binder maintained at the site pharmacy, indicating whether the subject should receive DFO or placebo.

8. BLINDING

In the phase I study, the reconstituted solution of DFO was colorless, and there were no specific treatment-related changes in laboratory tests or adverse events to suggest that the active drug can be identified from placebo. The study subject and all site staff, with the exception of the pharmacist, are blinded to the treatment assignment of all subjects. The pharmacists are specifically instructed not to reveal treatment assignment to the investigators. The SDMC staff (with the exception of the partially unblinded statistician, biostatistical programmer, and the senior programmer) are blinded to the treatment assignment of all subjects.

There is no specific antidote to DFO; therefore, unblinding is unnecessary in most cases. In cases of extreme emergency when the treating physicians request unblinding of treatment assignments for therapeutic purposes, the unblinding will only be revealed to the treating physicians but not the investigators. The treating physicians must be instructed not to reveal the blind to the subjects or study investigators. The study personnel (pharmacist) will be required to inform the Principal Investigator and the Project Manager within 24h in the event of unblinding. In cases where the treating physician is one of the study investigators, he/she also will be required not to reveal the identity of the study drug to other members of the study team, and not to perform subsequent study-related outcome assessments.

9. MULTIPLICITY

The primary futility hypothesis is tested via generalized linear model relating the probability of a favorable outcome to the treatment, as described in Section 11. The hypothesis is tested at the one-sided alpha level of 0.10.

Secondary outcomes analyses are considered exploratory. We present descriptive statistics and confidence intervals rather than p-values. The results of these analyses are interpreted with caution. Because the tests are not necessarily powered to detect a pre-specified effect size, a negative test result may be due to lack of power.

For safety outcomes, we will not account for multiplicity.

10. MISSING DATA

Based on previous experiences with clinical trials of acute stroke (including HI-DEF), it is anticipated that there would be minimal missing data for the 90-day assessment of the primary outcome. Only 1 of 42 HI-DEF subjects (2.38%) was missing the primary outcome. All efforts should be put forth to ensure near complete follow-up, in particular with the assessment of the primary outcome and occurrence of death. Nevertheless, minimal missing data may be inevitable.

If the missing data are minimal (<5%), we plan to perform a complete case analysis, wherein any subjects missing the primary outcome are excluded from the analysis. Otherwise, standard multiple imputation methods are used to account for missing primary outcome data in the analyses. A distribution for the primary outcome is derived from a logistic regression model that accounts for baseline covariates (the individual components of the ICH score (age, size and location of hemorrhage, Glasgow Coma Scale score, and presence of intraventricular hemorrhage), continuous time from onset to treatment initiation, serum glucose, statin use at onset), treatment, and post-treatment data (the last mRS obtained prior to end of study, initiation of anti-edema therapy), and a random sample from this distribution is used to impute values for missing primary outcomes. Multiple sample data sets with complete 90-day outcome are generated through PROC MI, and each of the data sets are analyzed according to the method described in Section 11.1.2, and the results (regression parameter and covariance matrix estimates) for each sample are combined and analyzed with PROC MIANALYZE to derive a valid statistical inference about the treatment effect.

As a sensitivity analysis, we plan to impute the missing primary outcome data via the Last Observation Carried Forward (LOCF) approach as well as the best/worst imputation. If the treatment effect is robust, we expect analyses using these imputation methods would yield similar inferences, particularly if the missing data are minimal (<5%).

11. PRIMARY ANALYSIS

11.1. Primary Outcome Variable Analysis

The primary outcome measure is the mRS score, dichotomized to define good functional outcome as a score of 0, 1, or 2 on the mRS at 90 days post randomization. For the primary analysis, mRS data collected no sooner than 60 days and no later than 120 days from randomization are considered for the primary outcome.

The primary futility hypothesis is tested via generalized linear model relating the probability of a good outcome to the treatment, after adjustments as described below.

The primary analysis is conducted under the modified ITT principle; all randomized subjects in whom the study infusion is initiated are included in the analysis and analyzed according to the randomly assigned treatment arm. Covariate misclassifications, however, will be corrected in the analysis in order to maintain the intended power (Fan L et al, 2015). OTT times greater than 24 hours will be included in the 12-24 hour category.

11.1.1. Statistical Hypothesis

The analysis model for the primary futility analysis is:

$$g\{P(Y_i = 1|\mathbf{x})\} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 ,,$$

where Y is the primary outcome variable (1=favorable, 0=unfavorable), $\mathbf{x}$ is a specified vector of covariates, and g is the link function. Here we consider $x_1 = 0$ if treatment=placebo and $x_1 = 1$ if treatment=DFO; $x_2, x_3, x_4, x_5, \text{and } x_6$ are indicator variables, respectively defining ICH score (0-2 vs 3-5), OTT time (≤12 vs 12-24 hours), baseline ICH volume (≤10 vs >10), screening NIHSS score (≤10 vs >10), and warfarin use (1=Yes, 0=No). Using this model, the primary statistical hypothesis is $H_0 : \beta_1 \geq 0.12$ versus $H_1 : \beta_1 < 0.12$.

11.1.2. Primary Analysis

The primary futility hypothesis is tested via generalized linear model relating the probability of a good outcome to the treatment, after adjustments as described above. This model is fit using the binomial distribution for Y and the identity link to derive an estimate of β_1, which yields an estimate of the risk difference for good outcome for DFO versus placebo. The PROC GENMOD procedure of SAS is used to obtain the test statistics and the results. If the model will not converge, the log link (to derive an estimate of the relative risk) and/or the logit link (to derive an estimate of the odds ratio) will be considered as alternative approaches.

If insufficient distribution of subjects in each level of anticoagulant use or ICH score results in quasi-complete separation of data points, these covariates will be excluded from the model. The ICH score is derived by assigning a point value to each of 5 components: the Glasgow Coma Scale score, ICH volume, presence of intraventricular hemorrhage, infratentorial origin, and age of subject. We do not anticipate significant collinearity between the ICH score and any individual component. However, if significant collinearity exists between the ICH score and ICH volume, the component adjustment will be excluded from the model.

The futility analysis will be conducted using a one-sided 90% upper confidence bound on the risk difference (the β_1 parameter), which is consistent with the one-sided alternative hypothesis and stated level of significance. To declare futility, the entire interval must lie below the value 0.12, indicating that the true difference in risk of good outcome is less than 0.12 with 90% confidence.

11.1.3. Exploring a Differential Treatment Effect According to OTT Window

The generalized linear model described above will be expanded to include an interaction between treatment and OTT window. While the trial will be underpowered to definitively address this question, the magnitude of the treatment effect, and corresponding confidence interval, will be estimated for each time window. The magnitude of the estimated differential effect can be used in determining whether future Phase III trials should maintain the 0-24 hour treatment window or consider limiting treatment to the ≤12 hour window. If the model does not converge, the simplified model described in Section 11.1.5 will instead be applied.

11.1.4. Adjusting for Covariates

The mRS is additionally assessed for treatment differences adjusting for a variety of pre-randomization covariates deemed clinically or prognostically important, such as age (continuous), location of hemorrhage (lobar vs. deep thalamic vs deep non-thalamic), screening Glasgow Coma Scale score (<8 vs 9-12 vs 13-15), ICH volume, IVH volume, serum glucose. Each covariate is evaluated individually first for a relationship with outcome. Each covariate which is found to be significantly associated with outcome will be added to the primary model in order to evaluate the impact of its inclusion on the estimated treatment effect. Covariates which cause more than a 20% change in the treatment effect estimate will be considered potential confounders, and a multivariable model that includes these potential confounders may then be constructed. If any components of the ICH score are identified as treatment modifiers, the ICH score will be removed from the multivariable model in favor of its individual components.

11.1.5. Subgroup Analyses

The primary outcome variable is assessed for treatment differences in the covariate categories listed below, assuming sufficient numbers of subjects (at least 60) are enrolled in each category. Each covariate is evaluated individually with a model that includes treatment, subgroup, and an interaction effect between the treatment (DFO or placebo). This model will be used to derive subgroup specific treatment effect estimates and the corresponding 95% confidence intervals, and the significance of the interaction presented to reflect heterogeneity across the subgroups.

- OTT time (≤12 vs >12 hours)
- ICH score: 0-2 vs 3-5
- Sex
- Race
- Ethnicity
- Age (≥60 vs <60)
- ICH volume (<30 vs ≥30 cm^3)
- Intraventricular hemorrhage (present vs absent)

11.1.6. mRS Dichotomized 0-3 v 4-6

Although mRS 0-3 is less favorable than the primary outcome of mRS 0-2, it would still be a desirable effect in patients with ICH given that no treatments exist to reduce disability. The trial is adequately powered to assess the futility hypothesis using mRS

0-3 as the outcome based on an absolute difference in treatment effect ≥13% in favor of DFO. This outcome will be analyzed as described above.

11.1.7. Ordinal Analysis of the mRS

As a strictly supportive analysis, a shift analysis of the full distribution of the mRS will be analyzed via proportional odds model, if and only if the proportional odds assumption holds. Given the total sample size, it may be the case that the cell count is insufficient to fit the full model with treatment and all covariates outlined in Section 11.1.1. If this is the case, a sequential approach will be explored, wherein ICH score and OTT window strata are individually added to a model that contains only treatment. The proportionality assumption will be assessed; the partial proportional odds model and/or the adjacent categories logit model may be considered depending on the type and degree of violation.

11.1.8. Per Protocol Analysis

Due to the nature of the study, variability in patient management is, to some extent, unavoidable. To account for this variability, a complete case analysis of the primary outcome variable is also conducted on a per protocol basis. The evaluable sample for this analysis includes all subjects who have at least one post-treatment assessment and no major protocol violations that affect the analysis. Subjects will be analyzed in the treatment arm representing the treatment actually received. Details for defining the per protocol sample were outlined by the Executive Committee (EC) prior to unblinding.

The per protocol analysis is based on the modified ITT population and excludes subjects according to the below criteria.

- Subjects in whom the qualifying ICH was due to a secondary cause will be excluded.
- Subjects in whom a major pre-specified eligibility criterion was violated.
    - Screening NIHSS <6 or screening GCS <6
    - Bilateral pulmonary infiltrates on screening CXR or other modifiers of ARDS risk specified in the protocol
- Subjects who did not receive all 3 doses (in full) of the study drug infusion
- Subjects in whom surgery (evacuation of the hematoma or hemicraniectomy) was performed before completing all 3 doses of the study drug infusion
- Subjects in whom withdrawal of care was instituted within 72 hours of enrollment.

11.1.9. ITT Analysis

Classically, ITT analysis is used in Phase III trials and includes all subjects randomized, regardless of whether they received the treatment or not. As a sensitivity analysis, the futility hypothesis will be assessed according to the full ITT principle. All subjects randomized will be included and considered in the treatment group to which he/she was randomized, regardless of the treatment actually received. The subjects in whom treatment is never initiated will be assigned an unfavorable outcome (mRS>2), based on the assumption that they did not receive

treatment due to clinical decline. Subjects in whom study drug was initiated but for whom the primary outcome is missing will be included via Last Observation Carried Forward.

11.1.10. As Treated Analysis

In the primary analysis, subjects are analyzed in the treatment arm to which they were randomly assigned, regardless of whether the correct treatment was administered. As a sensitivity analysis, the futility hypothesis will also be assessed in an as treated analysis. Subjects will be analyzed in the treatment arm representing the treatment actually received.

11.2. Secondary Outcomes Analyses

A series of secondary efficacy outcomes are evaluated. The list below outlines the planned analysis methods for the pre-specified secondary efficacy outcomes which are conducted with the SAS® Software System.

.
The usual verification of variable and model assumptions and goodness of fit assessments accompany each analysis. Some outcome measures may demonstrate a substantial departure from the normal distribution, even after transformation, in which case nonparametric methods may be considered.

These analyses are used to confirm or support the findings based on the primary outcome analysis. If most of the secondary outcomes show a change in the opposite direction from the primary or no change, we might have less confidence in the primary outcome.

11.2.1. Mortality at 90 days and at 180 days

The log-rank test is used to compare the survival curves for each treatment group (DFO versus placebo).

11.2.2. Symptomatic Cerebral Edema through 7 days or discharge

Chi-square test, or Fisher's exact test if the cell size is not sufficient, will be used to examine the effect of treatment on the incidence of symptomatic cerebral edema (neurological worsening/deterioration adverse events attributed to cerebral edema) up to day 7 or discharge, whichever is earlier.

11.2.3. Functional and Cognitive Measures at 90 days

Generalized linear modeling techniques are used to examine the effect of treatment on the dependent variables (NIHSS, MoCA, SIS-16) adjusting for previously specified covariates. As in the analysis of the primary outcome, data collected no sooner than 60 days and no later than 120 days from randomization are included. The NIHSS is a 42 point scale reflecting neurologic deficit. The MoCA is a 30 point scale assessing various domains of cognitive function. The SIS-16 is a 16 point scale, with total scores ranging from 16 to 80. These will be treated as continuous measures.

11.2.4. Analysis of the mRS at 180 days

The analyses specified in Section 11.1 will also be performed on the mRS at 180 days. Data collected no sooner than 150 days and no later than 210 days from randomization are included.

11.3. Exploratory Analyses

A series of exploratory analyses are also planned. The list below outlines the planned analysis methods for these pre-specified exploratory analyses which are conducted with the SAS® Software System.

The usual verification of variable and model assumptions and goodness of fit assessments accompany each analysis.

11.3.1. Treatment Effect on PHE Volume Progression

The effect of DFO treatment on PHE volume progression will be explored as a potential marker of biological activity on brain tissue. Measures of PHE volume progression will include absolute PHE volume, relative PHE, edema extension distance (EED) and the rate of PHE growth, derived from pre and post treatment scans by the central reader. These outcomes will be analyzed via general linear model which includes the baseline ICH volume and screening glucose as covariates.

11.3.2. Effect of PHE Volume Progression on Clinical Outcome

Progression of PHE is a post-randomization characteristic and can therefore be considered an outcome in its own right. The effect of PHE progression on outcome will be evaluated separately in the treatment and control arms via generalized linear model. The model will include adjustments for screening glucose, anti-edema agents, and hemicraniectomy.

11.3.3. Treatment Effect on Ventricular Enlargement

The effect of DFO on the size of ventricular enlargement post infusion in patients with intraventricular extension of ICH, not requiring an external ventricular drain, as a potential marker of treatment utility in intraventricular hemorrhage, will be explored via general linear model which includes the baseline size as a covariate.

12. SAFETY ANALYSES

12.1. Safety Monitoring

The detailed guidelines for monitoring for safety by the Executive Committee, the Medical Monitor, and the DSMB are included in the protocol.

The review of safety data will focus on the following Adverse Events of Special Interest (AEOSI):

- Anaphylaxis (at any time point during study drug infusion)
- Hypotension (defined as a decrease in blood pressure requiring medical intervention at any time point during drug infusion that cannot be explained by other causes)
- Development of new and unexplained visual or auditory changes after initiating treatment with the study drug
- Respiratory compromise

12.2. Summary of Adverse Events and Serious Adverse Events

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, timing relative to randomization, severity, and relatedness to the study drug.

For the AEOSIs previously defined, the relative risks and their 95% confidence intervals are provided. The DSMB is alerted if the lower confidence limit on the relative risk approaches 1 from below.

At the end of the study, treatment differences in the cumulative incidences of mortality, as well as the AEOSIs, are evaluated via the relative risk and corresponding 95% confidence interval.

13. REFERENCES

Efron B. Forcing a sequential experiment to be balanced. *Biometrika* 1971; 58: 403-417.

Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials*, *3rd ed*. 1998; New York: Springer-Verlag.

Fan L, Yeatts SD, Wolf BJ, McClure LA, Selim M, Palesch YY. The impact of covariate misclassification using generalized linear regression under covariate-adaptive randomization. *Stat Methods Med Res*. 2015 Nov 23. pii: 0962280215616405. [Epub ahead of print]