Supplemental Appendix – Statistical Methods

1. BACKGROUND

This Statistical Analysis Plan (SAP) describes the statistical methods to be used for the clinical study ML29093. The SAP was developed using version 3 of the study protocol dated 08 Jan 2015, and was finalized and approved before database lock and unblinding.

2. DETERMINATION OF SAMPLE SIZE

The sample size was originally determined as follows.

A sample size of 856 is required in order to achieve 80% power in the primary analysis to detect effect size of 9% absolute difference in the proportion of patients with favorable outcomes [mRS of 0 or 1 at Day 90] between the alteplase and control arms. The above power consideration assumes a control proportion of 65% and Type I error of 0.025 (one-sided), and uses a group sequential design with one interim analysis for futility (non-binding), based on O'Brien and Fleming boundary, conducted after 50% of the anticipated sample size have completed 90-day follow-up assessments. EAST® was used in sample size calculation.

The intention-to-treat principle will be applied to the primary analysis and, therefore, to safeguard against dilution of the treatment effect associated with an approximate 5% non adherence rate (due to loss to follow-up, consent withdrawal, treatment crossovers and stroke mimics), the above sample size was inflated by a factor of 1.108 (5). Therefore, a total sample size of 948 was planned for this study.

The study enrollment was terminated early in December 2016, after 313 patients were randomized. The analysis will be carried out based on all available data from patients randomized prior to termination. With this sample size, power calculations show that the study will be substantially underpowered to conduct superiority tests (**Table 1**). Therefore, all efficacy analyses will include estimation of treatment effects with the associated two-sided 95% confidence intervals.

Table 1 Estimated Power for the Final Sample Size (N=313)			
Assumed Effect Size	7%	8%	9%
Statistical Power [a]	25%	31%	38%

[a] Based on the t-test for the observed proportions of mRS 0-1 at Day 90, using type I error of 0.025, one-sided.

Assumptions are based on a blinded evaluation of non-adherences in the Jan 2017 Development Safety Update Report (DSUR) datacut:

- Control response rate among confirmed strokes is 74%.
- 10% of randomized subjects are stroke mimics and 80% of those stroke mimics will have favorable outcome.
- 1% of subjects received concomitant tPA
- 7.4% subjects without Day 90 outcome (290 effective sample size)

3. ANALYSIS TIMING

One interim analysis of the primary efficacy outcome was planned for clear futility once 50% of the subjects were enrolled. Because of the early termination of the study and failure to achieve the target enrollment, the interim analysis was not performed.

4. STATISTICAL METHODS

4.1. ANALYSIS POPULATIONS

4.1.1. Intent-to-Treat Population

All randomized patients are included in the intent-to-treat (ITT) population to adhere to the ITT principle. For efficacy analyses, treatment groups will be assigned as randomized (IV Alteplase 0.9 mg/kg + Aspirin Placebo, IV Alteplase Placebo + Aspirin 325 mg).

4.1.1 Safety Population

The Safety population is defined as all patients who received any amount of study drug. For safety analyses, treatment groups will be assigned according to the IV alteplase/alteplase placebo actually received. Thus, patients who receive active IV alteplase will be assigned to IV Alteplase 0.9 mg/kg + Aspirin Placebo group, and patients who receive IV alteplase placebo or did not receive the IV study drug (very rare) will be assigned to IV Alteplase Placebo + Aspirin 325 mg group. Group assignment will not take into account the actual dose of aspirin.

4.1 EFFICACY ANALYSES

All efficacy analyses will be to estimate treatment effects and provide the associated two-sided 95% confidence intervals.

Endpoints based on mRS will generally use imputation to achieve the complete ITT set, with the exception of the repeated measures modeling. Other efficacy variables will use all available data in the ITT population.

Assessment of efficacy will be conducted based on endpoints collected during the Day 90 visit. The protocol allows this visit to be scheduled within 14 days of the target 90 days post-treatment; however, a wider window of 30 days will be adopted for the purpose of efficacy analysis.

4.4.1 Primary Efficacy Endpoint

The primary efficacy analysis will estimate the effect of treatment with IV alteplase therapy and with the standard medical care (aspirin) in AIS patients with mild symptoms. The primary efficacy outcome is derived from the mRS.

The mRS is an assessment of disability with values from 0 (no symptoms at all) to 5 (severe disability); death will be scored as a '6'. The mRS is collected at 30 and 90 days after start of study drug administration. If death occurs prior to an assessment day (Day 30 or Day 90), the mRS will be considered available and will be set to 6.

The proportion of patients within each treatment group with a favorable outcome, defined by mRS of 0 or 1 at 90 days post-randomization, will be estimated in the IV alteplase arm and in the standard medical care arm. Reasons for not performing the assessment of mRS at Day 90 will be summarized.

The difference in proportions between the two groups will be estimated using the risk difference. The risk difference will be obtained from a linear model with the binary mRS 0-1 outcome as the response, treatment, and age, time from last known well to treatment, and baseline NIHSS as continuous covariates. Quadratic terms for the continuous covariates may be added to the model, if the Wald p-value for the quadratic term is <0.1. No interactions among the covariates will be included in this model.

A common approach for implementation of this model is binary regression, a GLM with the Bernoulli distribution for the outcome and the identity link function. This model has a possibility of non-convergence when the estimated response probability falls outside of the (0,1) interval. To circumvent the non-convergence problem, an approach of using the ordinary least squares with a robust standard error is used in this study. It was shown that if a robust standard error is used with the ordinary least squares estimation, the estimate of the risk difference is equal or similar to the estimate from the GLM with an identity link for Bernoulli outcome approach (22, 23).

The primary efficacy analysis will include all ITT patients, with patients grouped according to the treatment assigned at randomization. Missing or out of window mRS at Day 90 will be imputed with LOCF or late Day 90 mRS, if available, or simulated with a hot deck method (see Section 4.7 for handling of missing mRS).

As a supportive analysis of the primary efficacy endpoint, unadjusted or crude risk difference will be estimated along with the two-sided 95% CI using normal approximation. The standard error for the unadjusted risk difference will be derived from the group standard errors, assuming independence:

$$SE(D) = \sqrt{\frac{p(1-p)}{N_1} + \frac{q(1-q)}{N_2}}$$

where p and q are the observed group proportions and N_1 and N_2 are the group sample sizes.

The treatment effect for the binary mRS outcome will also be evaluated using the odds ratio statistic, derived from either a logistic regression, or a repeated measures model. These are considered exploratory analyses and described in Section 4.4.3.1.

4.4.2 Secondary Efficacy Endpoints

Secondary efficacy outcomes include mRS (full scale) and global favorable recovery at Day 90. These results will further describe the treatment effect.

4.4.2.1 Ordinal Modified Rankin Scale

As a secondary endpoint, mRS at Day 90 will be further explored on the range of the values (0 to 6) where the values of 5 and 6 are grouped. The distribution of full range

of mRS scores at Day 90 (imputed) will be summarized by treatment group. Both tabular displays and stacked bar charts will be prepared.

The comparison will be achieved by fitting a proportional odds model with mRS score at Day 90 as the dependent variable, and treatment group, pre-treatment NIHSS score, age, and last known well time to treatment as the continuous predictors (1, 8). Quadratic terms for the continuous covariates may be added to the model, if the Wald p-value for the quadratic term is <0.1. No interactions among the covariates will be included in this model. The adjusted odds ratio with its two-sided 95% CI will be shown. Direction of the odds ratio will be chosen to reflect change in the odds of achieving a lower (better) mRS if treated with IV alteplase.

The proportional odds assumption will be tested by the score test, and the p-value will be reported. If the proportional odds assumption does not hold, the resulting odds ratio will not be presented. The alternative analysis planned for this situation will provide the adjusted mean difference between mRS scores with two-sided 95% CI. Methodology for this estimation is similar to the primary outcome, but using the ordinal mRS as the dependent variable.

4.4.2.2 Global Outcome Measure

Global outcome measure is derived from mRS = 0 or 1, NIHSS = 0 or 1, BI ≥ 95, and GOS = 1 at Day 90, considered as a whole. If death occurs prior to Day 90, all four binary outcomes will be considered as not being achieved in the analysis.

The NIHSS score is the total from scoring 11 items, assessed by a clinician. Items that are not applicable are not included in the total. It is not expected that any items will be applicable but missing; in such cases, the total score will be missing. The total score ranges from 0 to 42. NIHSS score will be collected pre-treatment, and then after 22-36 hours from start of study drug administration, after 5 days or at discharge from hospital, and at Day 90.

The Barthel index (BI) measures activities of daily living by having an assessor rate 10 items, with possible scores of 0, 5, 10 or 15. A total score, ranging from 0 to 100, is computed as the sum of the ratings from the 10 items. As this is a rater-completed assessment, no missing data are expected. However, if any item is not scored, then a score of 0 will be given as it will be presumed the subject did not meet the defined criteria for the item (20).

Glasgow Outcome Scale (GOS) is an evaluation of the outcome of stroke, assessed on a scale ranging from 1 to 5 by a clinician. The scale will be reversed from the original published version (7) to assign lower scores to more favorable outcomes. The score is collected once in the study at Day 90.

The observed proportion of patients with a favorable outcome within each treatment group will be provided for the NIHSS (=0 or 1), BI (≥95) and GOS (=1) components of the global outcome measure, in a similar fashion as described for the primary outcome.

Treatment comparisons will be performed considering the four binary outcome measures as 4 dimensions of the outcome endpoint (12). These dimensions will be assumed to have a common odds ratio adjusted for pre-treatment NIHSS score, age, and last known well time to treatment. Quadratic terms for the continuous covariates may be added to the model, if the Wald p-value for the quadratic term is <0.1. No interactions among the covariates will be included in this model. Estimation will be performed using generalized estimating equations for a logistic model with correlation between the four outcomes of the same patient. The estimated common odds ratio and its two-sided 95% CI will be reported.

Each individual outcome measure will be evaluated separately: odds ratios with two-sided 95% CIs will be derived from a logistic model with the same covariates as used for the global outcome measure to facilitate interpretation of the results.

LOCF and hot deck imputation will be used for the missing mRS component, similarly to the primary analysis. No imputation is planned for the other missing components, other than the non-achievement imputation in the event of death. However, incomplete component sets can still contribute to the model and will be included.

4.4.3 Exploratory Efficacy Endpoints

4.4.3.1 Exploratory Analyses for the Primary Endpoint

The primary analysis of the mRS 0-1 outcome will be complemented with exploratory supportive evaluations of treatment group differences.

The following analyses will be conducted to construct risk differences for mRS 0-1. Missing values for day 90 mRS will be imputed in the same manner as for the primary efficacy analysis unless otherwise specified.

- Unadjusted risk difference.
- Adjusted risk difference, from a linear regression with treatment, age, onset time to treatment, baseline NIHSS, and propensity score as covariates. The propensity score is derived from a logistic regression for treatment group with imbalanced covariates included in the model as covariates. This analysis will be triggered by presence of baseline imbalance.

The following analyses will be conducted to construct the odds ratio for mRS 0-1 such that the direction of the odds ratio reflects change in the likelihood of achieving a favorable outcome if treated with IV alteplase:

- Logistic regression, adjusting for pre-treatment NIHSS score, age, and last known well time to treatment as continuous covariates. Quadratic terms for the continuous covariates may be added to the model, if the Wald p-value for the quadratic term is <0.1. No interactions will be included in this model.
- Repeated measures model
  mRS outcomes at Day 30 and Day 90 will be modeled jointly by a logistic regression, using the GEE approach to account for the correlation between Day 30 and Day 90 outcomes of the same subject. No imputation will be performed for missing mRS for this particular analysis.

4.4.3.2 Additional Exploratory Analyses

To characterize treatment response among the subset of subjects with confirmed strokes, adjusted risk difference for the primary efficacy outcome will be repeated on the subset of patients with stroke mimics excluded. In addition, analyses as specified in section 4.4.2 for the secondary efficacy outcome measures will be produced for confirmed strokes.

4.4.3.3 Modified Rankin Scale

The mRS score will be used to evaluate the proportion of patients with a score = 0 vs 1 or higher. Observed percentages and adjusted risk difference will be presented. The adjustment will be performed in the same way as for the primary endpoint. Adjusted odds ratio will be presented in a separate analysis for this outcome in the same fashion as for mRS 0-1.

4.4.3.4 National Institutes of Health Stroke Scale

Descriptive statistics (n, mean, standard deviation, median, quartiles, minimum, and maximum) will be provided for NIHSS score by treatment group, both at Day 5 / Hospital discharge and at Day 90. Reasons for not performing the assessment of NIHSS at Day 90 will be summarized.

The NIHSS score will be used to evaluate the proportion of patients with a score = 0 vs 1 or higher, and separately, 0-1 vs. 2 or higher. Observed percentages and adjusted risk difference will be presented. The adjustment will be performed in the same way as for the primary endpoint. Adjusted odds ratio will be presented in a separate analysis for this outcome in the same fashion as for mRS 0-1.

4.4.3.5 Barthel Index

The BI total score will be used to evaluate the proportion of patients with complete independence (defined as a score = 100) and will be summarized by treatment group. Observed percentages and adjusted risk difference will be presented. The adjustment will be performed in the same way as for the primary endpoint.Reasons for not performing the assessment of BI will be summarized.

4.4.3.6 Glasgow Outcome Scale

Raw GOS score will be summarized by treatment group, and include the percentage of patients having each GOS score, and the proportion of patients with good recovery. Observed percentages and adjusted risk difference will be presented. The adjustment will be performed in the same way as for the primary endpoint. Reasons for not performing the assessment of GOS will be summarized.

4.4.4 Subgroup Analyses

For the primary endpoint, consistency of treatment effect of alteplase will be evaluated by various pre-specified baseline covariates. The primary pre-specified variables and their analytic thresholds are defined by:

- Age group (< 65 vs. ≥ 65)
- Pre-treatment NIHSS score group (0-2 vs. 3-5)
- Last known well time to treatment group (0-2 hours vs. > 2 hours)
- Stroke subgroup (RISS vs. non-RISS)

Adjusted odds ratios by subgroup will be obtained from multivariable logistic regressions with the main effects for treatment, age, last known well time to treatment, and NIHSS, and fit separately for each level of the subgroup.

Interactions of IV alteplase treatment effect with each of the subgroup variables will be explored by adding interactions of the subgroup variables with treatment to the logistic regression. The interactions will be included one at a time; thus, there will be 4 separate models to test each interaction. Results from these models will not be reported, except when the interaction p-values is smaller than 0.1, in which case the p-values will be footnoted in the forest plot.

A forest plot will be constructed to visually illustrate the subgroup analyses. The plot will show the odds ratios described above with two-sided 95% CIs, corresponding to each level of these subgroups. The plot will also show the common odds ratio estimate in different color or line pattern, as a reference. A subgroup analysis will be conducted only if there is at least 10 subjects per outcome (mRS 0-1, mRS >1), and the standard error of the odds ratio estimate is stable with the addition of the covariates.

4.5 COGNITIVE, BEHAVIORAL AND PATIENT REPORTED OUTCOME ANALYSES

All the cognitive, behavioral and patient reported outcomes are considered exploratory.

4.5.1 Cognitive Assessments

A series of tests will be conducted at Day 90 to evaluate cognitive abilities of the patients. The tests and their outcomes are presented in the following table:

Table 2
Cognitive Assessments

Test	Outcome
Controlled Oral Word Association test	Number of words listed
Hopkins Verbal Learning test	Number of words recalled in 3 trials, Retention, and Recognition Discrimination Index
Digit symbol coding from the Wechsler Adult Intelligence Scale	Number of correctly coded digits
Forward and Backward Digit Span test	Maximum number of digits repeated, forward and backward scores
Benton Judgment of Line Orientation, form V	Total score
Animal Naming test	Number of unique animals named
15-item Boston Naming test	Total number of items answered correctly either spontaneously or after a stimulus cue

Outcomes from each test will be summarized by treatment group (mean, standard deviation, median, quartiles, minimum and maximum). Mean difference between treatment groups with 95% unadjusted normal CI will be presented. Reasons for the assessment not being performed will be summarized by the number and percentage in each category.

4.5.2 Ambulatory Performance

The 10 meter walk test will be conducted at Day 90 visit to measure patient's walking speed. The speed will be measured at a comfortable pace and a fast pace, giving two attempts for each pace level. Average speed from the two attempts obtained at each pace will be summarized descriptively by treatment group. If speed can only be calculated in one of the attempts at a pace level, this result will be used.

Average speeds will be summarized by treatment group (mean, standard deviation, median, quartiles, minimum and maximum). Mean difference between treatment groups with 95% unadjusted normal CI will be presented. Reasons for the assessment not being performed will be summarized by the number and percentage in each category.

4.5.3 Patient Reported Outcomes

The patient reported outcomes (PROs) for this study include the EQ-5D, the 16 item Stroke Impact Scale (SIS-16), and the Center for Epidemiologic Studies Depression Scale (CES-D).

4.5.3.1 EQ-5D

Health-related quality of life will be assessed using the EQ-5D. The EQ-5D is a generic, preference-based health utility measure with questions about mobility, self-care, usual activities, pain/discomfort, and anxiety/depression that are used to build a composite of the patient's health status. The second component of the EQ-5D is a visual analogue scale (VAS), asking patients to rate their health from 0 to 100 where 0 represents the worst imaginable health state and 100 represents the best imaginable health state.

Each question is scored as 1 for no problem, 2 as some problem, and 3 indicating extreme problem. Ambiguous values (eg. 2 boxes are ticked for a single dimension) will be treated as missing (neither value used).

The 5 scores will then be converted into a single summary index by applying the time trade-off valuation technique for the US, published by EuroQol group. The resulting summary index is a continuous variable ranging from -0.11 (i.e., 3 for all questions) to 1.0 (i.e., 1 for all questions) on a scale where 0.0 = death and 1.0 = perfect health (18). If any of the 5 scores are missing, the summary index will not be calculated and set to missing.

The VAS is scored by the crossing point of the VAS with the response line, or with a horizontal line extended from the end point of the response line. Ambiguous values (eg. the line crosses the VAS twice) will be treated as missing.

Scores to each question, the summary index, and the VAS result, will be summarized descriptively, and the two-sided 95% CI of the mean difference will be presented for the index and VAS.

4.5.3.2 Stroke Impact Scale-16 (SIS-16)

Physical functioning will be evaluated using the SIS-16, which is a validated, stroke-specific, quality of life measure to assess the impact of stroke on a patient's health and life. Each of the 16 items is rated using the following scale: 1 = could not do at all; 2 = very difficult; 3 = somewhat difficult; 4 = a little difficult; 5 = not difficult at all.

The physical domain score is obtained by averaging the non-missing responses, provided that 9 or more questions are answered. The score is calculated as:

$$\text{Score} = 100 \times [(\text{Mean} - 1) / (5 - 1)]$$

which ranges from 0 to 100 (19).

Responses to individual questions and the physical domain score will be listed. The domain score will be summarized by treatment group and the two-sided 95% CI of the mean difference will be presented.

4.5.3.3 Center for Epidemiologic Studies Depression Scale (CES-D)

The CES-D is a short, self-report scale designed to measure depressive symptomatology. Responses to the 20 individual items will be scored to obtain an overall depression assessment, as follows.

1. Assign scores 0 = rarely or none of the time, 1 = some or little of the time, 2 = occasionally or moderate amount of time, 3 = most or all of the time.
2. Reverse positive items by subtracting their score from 3. Positive items are 4, 8, 12, and 16.
3. If 4 or fewer items are missing, sum the available item scores to determine the total score. If more than 4 items are missing, no score is calculated.

The total score will be used to evaluate the proportion of patients with depressive symptoms (defined as a score of $\geq$16 at 90 days) and will be summarized by treatment group.

Observed percentages and adjusted odds ratio will be presented. The adjustment will be performed in the same way as for the exploratory analysis of mRS 0-1.

4.6 SAFETY ANALYSES

MedDRA (using the latest version at the time of reporting) will be used as the thesaurus for AEs and diseases, and the World Health Organization (WHO) drug dictionary will be used for treatments and MedDRA for procedures. A glossary of these codes will be produced.

AEs with onset prior to initiation of study treatment which do not worsen after study treatment or with onset after Day 90 assessment, will be listed as non-treatment emergent and will not be included in the safety summaries.

4.6.1 Exposure to Study Medication

Exposure data will be listed for both study drugs. The listing will show the randomized treatment as well as the treatment actually received.

A summary of exposure to alteplase/alteplase placebo will be produced for the safety population. The total dose administered and duration of infusion will be summarized by active or placebo Alteplase, as actually received. For Aspirin, dose received will always be “325 mg”, and therefore will not be summarized.

Compliance with study medications will be summarized for the safety population. For Alteplase, counts of infusion altered or stopped prematurely will be tabulated along with the reason. For aspirin, the number of patients who did not take study aspirin and a summary of reasons will be presented.

4.6.2 Adverse Events

AE safety endpoints in this study are:

- Number and percentage of patients experiencing serious and non-serious AEs after receiving study drug
- Number and percentage of patients experiencing serious and non-serious AEs of special interest

AEs will be coded (using the latest MedDRA version at the time of reporting) and tabulated by system organ class (SOC); individual events within each SOC coded as preferred terms will be presented in descending frequency using the safety population. Treatment groups will be summarized according to the actual treatment received.

AEs will also be tabulated by intensity (severity) and relationship to study drug, as indicated by the investigator. Intensity of AEs will be graded according to the WHO Toxicity Grading Scale for Determining the Severity of Adverse Events. Incidence and severity of AEs will be described for all treated patients by treatment group and by the MedDRA classification.

A summary of incidence rates of all deaths and stroke-related or neurological deaths will be presented. Cumulative incidence of all deaths and stroke-related or neurological deaths will be presented by means of Kaplan-Meier plots if there are sufficient events (at least 5 in each group).

Timing of the AEs relative to Day 30 and Day 90 will be determined, using imputation for partial AE start and stop dates if necessary. AEs occurring after receipt of any study medication and before Day 90 will be summarized. AEs leading to IV study drug modification will be listed. SAEs, ICH, AEs of special interest and AEs leading to study treatment discontinuation will be summarized through Day 90. In addition, non-serious AEs within 30 days will be summarized.

4.6.2.2 Adverse Events of Special Interest

Non-serious AESIs are required to be reported by the investigator to the Sponsor immediately (i.e., no more than 24 hours after learning of the event; see Section 5.4.2 of protocol for reporting instructions).

AESIs for this study include the following:

- sICH events, if not already reported as an SAE
- Stroke recurrence, if not already reported as an SAE.
  Stroke recurrence is defined as a new ischemic stroke based on the AE reporting. Stroke events will be identified through a medical review of the AE preferred terms.
- Suspected Transmission of an Infectious Agent via a Medicinal Product (STIAMP) by the study drug, if not already reported as an SAE.

AESIs will be tabulated by SOC and preferred term within SOC by treatment group using the safety population.

4.6.3 Neuroimaging

Neuroimaging may consist of a non-contrast computed tomography (CT) or magnetic resonance imaging (MRI) of the brain. Non-contrast CT or MRI will be performed at baseline to ensure that the patient does not have evidence of acute intracranial hemorrhage (ICH) prior to study drug administration.

An additional non-contrast CT or MRI will be performed at 22 to 36 hours from the initiation (bolus) of study drug infusion. The purpose of this is to evaluate for ICH in the patient. Additional CT scans or MRIs should be performed based on the investigator's discretion, or if any clinical findings suggest ICH.

All CT and MRI scans will be evaluated by a central team of radiologists blinded to the clinical attributes of the case. Results from the evaluation of the CT or MRI scans by both local and central readings will be listed, as available.

4.6.3.1 Intracranial Hemorrhage

Intracranial hemorrhage events (ICH) are identified by neuroimaging and are recorded as AEs using one of the terms specified in protocol. The images obtained within the first 36 hours after treatment are further evaluated by central reading, which may or may not confirm ICH.

While ICH may be identified at any time during the study, the first 36 hours after treatment are the key time interval. If multiple scans indicate ICH within 36 hours, it is interpreted as a single ICH event. ICH events identified after the first 36 hours will contribute to the general AE summaries.

Incidence of ICH within 36 hours by AE reporting will be presented by treatment group. Determination of ICH within 36 hours from local and central readings of scans will also be presented, and agreement between the local and central readings will be tabulated.

Difference in percentage of subjects with ICH within 36 hours based on central and local scan reading will be reported by treatment group. A two-sided 95% CI around this difference will be constructed using the Miettinen-Nurminen score method (20).

Additionally, central reading classifications will be used to further describe subtypes of any ICH.

The central reading will classify ICH according to the following intracerebreal subtypes:

- Hemorrhagic infarct type 1 or type 2
- Parenchymal hematoma type 1 or type 2
- Remote intraparenchymal hemorrhage type 1 or type 2

In addition, subarachnoid hemorrhage, subdural hemorrhage, or epidural hemorrhage will be noted. ICH subtypes will be tabulated by treatment group.

4.6.3.2 Symptomatic Intracranial Hemorrhage

The protocol specifies that ICH is considered symptomatic (sICH) if it is not seen on CT or MRI scan at baseline, and any neurologic decline is attributed to it by the local investigator. Both serious sICH and non-serious sICH events are required to be reported by the investigator to the Sponsor immediately.

sICH within 36 hours is considered the primary safety outcome. sICH events are captured through AE reporting.

In addition to the primary definition of sICH, the following two definitions of sICH will be explored and reported in this study.

- sICH using Heidelberg classification: either (1) PH2 as reported by radiology core, or (2) an sICH as classified by PRISMS but also associated with 4-point total NIHSS score worsening or 2 point worsening on any single NIHSS item when comparing scores immediately pre and post scan.
- sICH using SITS-MOST definition (0): PH2 as listed by radiology core on CT scan between 22-36 hours, which is associated with deterioration of ≥4 points on the NIHSS from the lowest NIHSS value between baseline and 36 hours, but prior to repeat CT, or leading to death.

Percentage of subjects with sICH within 36 hours will be presented for each definition by treatment group. Difference between treatment group percentages will be reported along with a two-sided 95% CI constructed using the Miettinen-Nurminen score method (20).

ICH symptomatology will be tabulated for confirmed strokes and separately for stroke mimics.

For ICH and sICH events, the following rules will be used to determine their timing relative to 36 hours:

- For ICH determination by scan reading (local or central), take the earliest post-baseline scan with bleeding identified. Use the evaluation time of that scan to compare to 36 hours.

- For ICH determination by AE reporting, take the start date of the corresponding AE. Then find all scans with the same date of local reading, where bleeding is identified. Take the earliest such scan and use the evaluation time of that scan to compare to 36 hours. If no scan is found, use 12:00 am on the AE start date to compare to 36 hours.

For sICH using SITS-MOST definition, when there are multiple scans associated with a PH2 bleed in a subject, the first of these scans will be used to select pre- and post-scan NIHSS scores.

4.7 HANDLING OF DROPOUTS, MISSING/INCOMPLETE DATA, OR OUTLIERS

For the analyses of the mRS based endpoints, last observation carried forward will be used to impute missing or out of window assessments of mRS at Day 90. The mRS result carried forward needs to be on or after Day 4 to qualify for this imputation. If there is no in-window or LOCF assessment of mRS, then a late Day 90 mRS will be used if available.

ITT patients without any observed mRS past day 4 will have their Day 90 mRS simulated by a hot deck method. This will be an adjustment cells imputation (24), with the cells defined by treatment group, age (< 65 vs. ≥ 65), pre-treatment NIHSS score (0 − 2 vs. 3 − 5), and last known well time to treatment (0 − 2 hours vs. > 2 hours). For each patient with a missing mRS score, a “donor” patient will be drawn at random from the pool of patients with available mRS and who belong to the same cell as the recepient (i.e. randomized to the same treatment group and having the same categories for age, last known well time to treatment and baseline NIHSS). The imputed mRS will then be the mRS value of the “donor”.

It is expected that only a small number of patients will need the hot deck imputation; therefore, the techniques recommended in (24) for a variance estimate that

incorporates the additional variance from the missing information will not be implemented.

No imputation of missing patient reported outcomes, cognitive assessments, or the 10-meter walk test will be performed except as specified in the scoring rules for the assessment. Reasons for the assessment not being performed will be summarized by the number and percentage in each category by treatment group.

5. REFERENCES

1. Austin P. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011 May; 46(3): 399–424.
2. Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J, and the Optimising Analysis of Stroke Trials (OAST) Collaboration. (2007) Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. Stroke; 38: 1911-1915.
3. Dhamoon MS, Moom YP, Boden-Albala B, et al. (2009) Long-term functional recovery after first ischemic stroke: the Northern Manhattan Study. Stroke; 40: 280-511.
4. Edlow JS, Smith EE, Stead LG, et al. (2013) Clinical policy: Use of t-PA for the management of acute ischemic stroke in the Emergency Department. Ann Emerg Med; 61: 225-243.
5. Friedman LM, Furberg CD, DeMets DL. (1998) Fundamentals of Clinical Trials. Third ed., New York, NY: Springer: 108.
6. Jauch EC, Saver JL, Adams HP, et al. (2013) Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke; 44: 870-947.
7. Jennett B., and Bond M. (1975) Assessment of outcome after severe brain damage. Lancet; 1(7905): 480-484.
8. McHugh GS, Butcher I, Steyerberg EW, et al. (2010) A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT Project. Clin Trials; 7: 44-57.
9. Pampallona S, Tsiatis AA, Kim K. (2001) Interim monitoring of group sequential trials using spending functions for the type I and type II error probabilities. Drug Inf J; 35: 1113-1121.
10. Reeves M, Khoury J, Alwell K, et al. (2013) Distribution of NIHSS in the Cincinnati/Northern Kentucky Stroke Study. Stroke; In press.

11. Rubin D. (1987) Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons

12. Tilley BC, Marler J, Geller NL, et al. (1996) Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA Stroke Trial. Stroke; 27:2136-2142.

13. Ulm K. (1990) A simple method to calculate the confidence interval of a standardized mortality ratio. Am J Epidemiol; 131: 373-375.

14. Wei LJ (1978) The Adaptive Biased Coin Design for Sequential Experiments. The Annals of Statistics; 6(1): 92-100.

15. Efron B. (1971) Forcing sequential experiments to be balanced, Biometrika, 58, 403-417.

16. Zhao W., Ciolino J., and Palesch Y. (2010) Step-forward Randomization in Multicenter Emergency Treatment Clinical Trials. Academic Emergency Medicine; 17(6): 659-665.

17. Zhao W., Weng Y. (2011) Block urn design — A new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. Contemporary clinical trials; 32(6): 953-961.

18. Calculating the U.S. Population-based EQ-5D Index Score: Research Initiative in Clinical Economics. February 2005. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/professionals/clinicians-providers/resources/rice/EQ5Dscore.html (last accessed on 8 October 2014).

19. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The Stroke Impact Scale version 2.0. Evaluation of reliability, validity, and sensitivity to change. Stroke 1999; 30: 2131-2140.

20. Mahoney FI, Barthel D. Functional evaluation: the Barthel Index. Maryland State Medical Journal 1965;14: 56-61.

21. Liu GF, Wang J, Liu K, Snavely DB. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Statistics in Medicine 2006; 25(8): 1275-86.

22. Mukaka M, White SA, Mwapasa V, Kalilani-Phiri L, Terlouw DJ, Faragher EB Model choices to obtain adjusted risk difference estimates from a binomial regression model with convergence problems: An assessment of methods of adjusted risk difference estimation. Journal of Medical Statistics and Informatics 2016; 4:5.

23. Cheung YB. A modified least-squares regression approach to the estimation of risk difference. Am J Epidemiol. 2007; 166:1337-44.

24. Andridge RR, Little RJA. A Review of Hot Deck Imputation for Survey Non-response. Int Stat Rev. 2010 April; 78(1): 40–64.

Wahlgren N, Ahmed N, Dávalos A, Ford GA, Grond M, Hacke W, Hennerici MG, Kaste M, Kuelkens S, Larrue V, Lees KR, Roine RO, Soinne L, Toni D, Vanhooren G; SITS-

MOST investigators. Thrombolysis with alteplase for acute ischaemic stroke in the Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST): an observational study. Lancet 2007 Jan 27; 369 (9558): 275-82.