101SK202/NCT02730455

Statistical Analysis Plan

STATISTICAL ANALYSIS PLAN

Product Studied: Natalizumab
Protocol Number(s): 101SK202

A Multicenter, Double-Blind, Placebo-Controlled, Randomized, Parallel-Group, Dose-Ranging Study to Evaluate the Safety and Efficacy of Intravenous Natalizumab (BG00002) in Acute Ischemic Stroke

Version 2.0

Approved By:

[redacted], SMT Statistician — Date

[redacted]

[redacted], CDT Statistician — 22 NOV 2011 — Date

[redacted], CDT Medical Director — Date

STATISTICAL ANALYSIS PLAN

Product Studied: Natalizumab
Protocol Number(s): 101SK202

A Multicenter, Double-Blind, Placebo-Controlled, Randomized, Parallel-Group, Dose-Ranging Study to Evaluate the Safety and Efficacy of Intravenous Natalizumab (BG00002) in Acute Ischemic Stroke

Version 2.0

Approved By:

[redacted], SMT Statistician — Dec 05, 2017
Date

[redacted], CDT Statistician — Date

[redacted], CDT Medical Director — Dec 1, 2017
Date

Table of Contents

History of Changes

Date	Author	Change(s)
15Nov2017	[redacted]	Clarify the utility of Bayesian framework to characterize treatment effect.
15Nov2017	[redacted]	Additional analysis on the primary endpoint to pool ACTION and ACTION 2 data together to estimate overall treatment effect.
21Nov2017	[redacted]	Typos corrected.

1 Description of Objectives and Endpoints

Primary Objective

The primary objective of the study is to assess the clinical effects of natalizumab versus placebo in acute ischemic stroke on clinical measures of functional independence and activities of daily living.

Primary Endpoint

The primary efficacy endpoint is a composite global measure of functional disability based on a score of 0 or 1 on the mRS and a score of ≥95 on the BI at Day 90.

Secondary Objectives

The secondary objectives of this study are designed to explore dose and exposure response, time dependency of the treatment effect, and the clinical treatment effects of natalizumab versus placebo in acute ischemic stroke on the following measures of independence, activities of daily living, neurologic function, quality of life, cognition, and safety and tolerability.

Secondary Endpoints

The secondary endpoints are as follows:

- modified Rankin Scale (mRS) score at Day 90
- Barthel Index (BI) score at Day 90
- Stroke Impact Scale-16 (SIS-16) score at Day 90
- Montreal Cognitive Assessment (MoCA)score at Day 90
- Safety (incidence and proportion of Adverse Event (AE)s and Serious Adverse Event (SAE)s)
- National Institute of Health Stroke Scale (NIHSS) score at Day 90

Exploratory Endpoints

The additional/exploratory endpoints are as follows:

Exploratory objectives of this study are to evaluate the effect of natalizumab on measures of function, cognition, fatigue, depression, quality of life, and pharmacokinetic (PK) /pharmacodynamic (PD) relationships over time.

Exploratory Endpoints

The exploratory endpoints are as follows:

- Functional Independence Measure (FIM) score at Day 90
- Symbol-Digits Modalities Test (SDMT) score at Day 90
- Fatigue Severity Scale (FSS) score at Day 90
- Beck Depression Inventory 2 (BDI-2) score
- Serum concentrations of natalizumab at selected times after dosing
- Blood biomarkers of natalizumab
- Subject direct resource use (assessed using a health resource utilization [HRU] questionnaire)
- EuroQoL EQ-5D-3L (questionnaire)

2 Study Design

Study Overview

This is a Phase 2, proof-of-concept, multicenter, double-blind, placebo-controlled, randomized, dose-ranging study of natalizumab in subjects with acute ischemic stroke, administered at ≤ 24 hours from when the subjects were last known normal (LKN). This study will evaluate the efficacy and safety of natalizumab over a 90-day period at approximately 67 sites in the US and Europe.

Approximately 270 subjects may be randomized in the study. Of these, no more than 90 subjects will be treated between >9 and ≤24 hours from LKN, and the rest will be treated ≤9 hours from LKN. Randomization will occur separately within each treatment window. For subjects in the ≤9hour treatment window, randomization will be stratified by baseline NIHSS category (NIHSS scores from 5 to 15 or 16 to 23), tissue plasminogen activator (tPA) use (yes or no), and region; for subjects in the >9 to ≤24 hour window, randomization will be stratified by tPA use (yes or no) and region (USA, Spain, Germany and UK). Within each treatment window, they will be randomized in a 1:1:1 ratio to receive a single dose of 300 mg or 600 mg IV natalizumab or placebo.

After receiving their single dose of study drug, subjects will have post-treatment assessments performed at 12 ± 3 hours, 24 ± 6 hours, and Day 5 (or prior to discharge), Day 30± 5 days, and Day 90± 5 days.

Overall Study Duration and Follow-Up

The study will consist of Screening, Baseline, Randomization, Treatment, and Post-Treatment periods. The overall duration of participation for each subject in the study will be approximately 90 days.

Screening

Subject eligibility for the study will be determined at the time of acute ischemic stroke diagnosis.

Treatment

Eligible subjects will receive study treatment at ≤24 hours from their LKN. Approximately one-third of the subjects will be treated in each treatment group.

Post-Treatment and Follow-Up

Subjects will be assessed for follow-up at 12 ± 3 hours, 24 ± 6 hours, and Day 5 (or prior to discharge), Day 30± 5 days, and Day 90± 5 days after the start of study treatment administration. The Final Visit will be Day 90.

Study Stopping Rules

Biogen may terminate this study at any time, after informing Investigators. Investigators will be notified by Biogen or designee if the study is placed on hold, completed, or closed. There are no pre-specified stopping rules. An independent Data Safety Monitoring Committee (DSMC) was formed and is reviewing unblinded data regularly to assess safety and risk-benefit. Details of the DSMC responsibilities are provided in the DSMC charter.

End of Study

The end of study is last subject, last visit for final collection of data.

Table 1: Schedule of Activities for Study 101SK202 (protocol version 2.0)

Tests and Assessments	Screening[1,2]	0 Hours (Day 1)	Within 1 Hour After Start of Infusion	12 Hours ±3 Hours	24 Hours ±6 Hours	Day 5[3]	Day 30[4] ±5 Days	Day 90 Follow-up[4]/ Early Termination[4] ±5 Days
Informed consent	X							
Confirm eligibility	X							
Demographics and medical history[5]	X							
Physical and neurological examination	X				X	X	X	X
Vital signs[6]	X	X	X		X	X	X	X
Height (if available) and weight	X (anytime from Screening to Day 5 Visit)							
Urine pregnancy test[7]	X							
Hematology and blood chemistry	X			X	X	X	X	X
Serum biomarkers	X			X	X	X	X	X
PD sampling	X		X	X	X	X	X	X
PK sampling	X		X	X	X	X	X	X
Blood sample for anti-natalizumab antibodies	X						X	X
NIHSS	X	X[2]			X	X	X	X
mRS						X[8]	X	X
BI						X	X	X
SIS-16						X[9]	X	X
MoCA						X	X	X
Functional Independence Measure							X	X
SDMT						X	X	X
Fatigue Severity Scale						X[9]	X	X
Beck Depression Inventory 2						X[9]	X	X
HRU questionnaire						X	X	X
EQ-5D-3L						X	X	X
Study treatment administration[10]		X						
Stroke subtype classification[11]						X		
SAE reporting	Record as per Section 15.2 of the protocol							
AE reporting	Record as per Section 15.2 of the protocol							

Tests and Assessments	Screening[1,2]	0 Hours (Day 1)	Within 1 Hour After Start of Infusion	12 Hours ±3 Hours	24 Hours ±6 Hours	Day 5[3]	Day 30[4] ±5 Days	Day 90 Follow-up[4]/ Early Termination[4] ±5 Days
Concomitant medications	Record as per Section 11.5 of the protocol							
Concomitant procedures	Record as per Section 11.5 of the protocol							

AE = adverse event; BI = Barthel Index; BDI-2 = Beck Depression Inventory 2; CT= computed tomography; ECG = electrocardiogram; FIM = Functional Independence Measure; FSS = Fatigue Severity Scale; HRU = health resource utilization; LKN = last known normal; MoCA = Montreal Cognitive Assessment; MRI = magnetic resonance imaging; mRS = modified Rankin Scale; NIHSS = National Institute of Health Stroke Scale; PD = pharmacodynamic; PK = pharmacokinetic; SAE = serious adverse event; SDMT = Symbol-Digits Modalities Test; SIS-16 = Stroke Impact Scale-16.

Note: All timepoints are relative to start of study treatment administration.

1 Screening assessments that are performed as standard of care do not need to be repeated for the study, provided that Biogen has access to the information and data are recorded in subject's case report form. Data collected for the Screening assessments will be used for patient baseline analysis in this study. The ECG and CT or MRI assessments results should already be available at Screening (as they are performed as part of the standard of care), and a copy of the images and report should be filed with the subject's study records.

2 All Screening assessments must be performed prior to infusion. The 0 Hours vital signs assessments are to be performed within 15 minutes prior to study treatment administration. If the Screening NIHSS is performed more than 1 hour prior to study treatment administration, it must be repeated during the visit at 0 Hours before study treatment administration.

3 Day 5 assessments are to occur on Day 5 or earlier if discharged, but must occur prior to discharge.

4 The subject will be asked to complete the Day 30 and Day 90 Follow-Up or Early Termination assessments in person. If the subject is unable to return to the study center to complete the Day 30 or Day 90 (Follow-up or Early Termination) assessments in person, safety information will be collected by telephone or remotely, as local regulations allow and pending medical monitor approval, and will include the collection of AEs, SAEs, FIM, mRS, BI, BDI-2, SIS-16, EQ-5D-3L, physical/neurological exam, vital signs, hematology/blood chemistry, serum biomarkers, PD sampling, PK sampling, anti-natalizumab antibodies, NIHSS, MoCA, SDMT, FSS, HRU, and concomitant medications.

5 Medical history should include an assessment of prior substance abuse.

6 Vital signs collected at each of the specified timepoints include temperature, blood pressure, pulse or heart rate, and respiratory rate. Vital signs collected as part of the subject's standard of care and that fall within 30 minutes of a study visit do not need to be repeated for the study, provided that Biogen has access to the information and data are recorded in the subject's case report form.

7 Required only for women of childbearing potential. If a serum beta human chorionic gonadotropin test was performed as part of the subject's standard of care, the results will be accepted in lieu of the urine pregnancy test.

8 An mRS assessment will be done on Day 5, or earlier if discharged.

9 Per instructions on the SIS-16, BDI-2, and FSS forms, the subject or proxy should provide responses as it applies to the prior 2 week period. When completing the SIS-16 BDI-2, and FSS assessments during the Day 5 Visit, the subject or proxy should consider only the time period starting between the initial stroke (LKN) and the Day 5 Visit.

10 The subject must be observed for 1 hour after completion of infusion.

11 The stroke etiology will be assessed based on the subject's standard of care and diagnostic testing. This should be completed at Day 5 or earlier if discharged, but if results are incomplete, it may be completed at Day 30.

3 Study Subjects

Descriptive analyses described in this section will be presented by randomized treatment group and for all study subjects combined in the intent-to-treat (ITT), unless otherwise specified. The ITT population is defined as those subjects who are randomized.

For categorical endpoints, number of subjects and percentage will be summarized. For continuous endpoints, summary statistics will be presented (n, mean, standard deviation, median, range). All summaries will be by treatment group and overall. Baseline is defined as the closest non-missing value prior to the infusion of study treatment. Subjects will be analyzed by the treatment to which they are randomized.

3.1 Subject Accountability

The number of subjects randomized and dosed during the randomized treatment period will be presented by country and site.

Disposition of the subjects will be summarized by treatment group to which they are randomized. The summary data will include: number (and percentage) of subjects randomized, number of subjects dosed, number withdrawing prior to dosing, number who received the entire infusion of study drug, number who completed the study, number who withdrew from study early, and the reasons for study withdrawal.

A subject listing of the reasons for study withdrawal will also be presented by randomized treatment group and subject.

3.2 Demography and Baseline Disease Characteristics

Demographic data, including age (years), age category <60, 60-69, 70-80), sex, ethnicity and race category, country, history of diabetes, glucose level on entry will be summarized. In addition, height (cm), weight (kg), and body mass index (BMI) (kg/m^2) captured during hospitalization will be presented.

Baseline disease characteristics will also be summarized, including type of modality used to assess stroke, stroke subtype and location of stroke.

Treatment time window is defined by the time from LKN to start time of study drug administration. The number of subjects in the treatment time will be summarized overall and by the following subcategories: <= 3 hours from LKN, > 3 to <= 4.5 hours from LKN, > 4.5 to <= 6 hours from LKN, > 6 to <= 7.5 hours from LKN, > 7.5 to <= 9 hours from LKN and > 9 to <= 12 hours from LKN, > 12 to <= 18 hours from LKN, > 18 to <= 24 hours from LKN and >24 hours from LKN (if applicable). Time from LKN to study drug infusion will be summarized by treatment group and overall using summary statistics.

Number and percentage of subjects who received any stroke therapies or stroke interventions (tPA -yes/ no) and (thrombectomy- yes/no) from the time of the current stroke onset to the infusion of study treatment will be summarized. Time from LKN to the start of tPA infusion (hours) and time from start of tPA infusion to start of study drug infusion will be summarized. Also the type of tPA administration method used (Intravenous（IV), Intra-arterial（IA), both IV and IA), mechanical thrombectomy and stenting use will be summarized as well.

The number and percentage of subjects for baseline NIHSS categories (NIHSS score 5-15, 6-23) will be summarized. The baseline NIHSS score will be summarized as well.

The above demographic and baseline characteristics will also be presented for the Modified intent-to-treat (MITT, defined in Section 6) and per-protocol (PP, defined in Section 6) populations.

Medical history and tobacco, alcohol and drug use will be summarized by treatment group and overall.

Demographics and baseline disease characteristics will also be summarized by treatment window (≤9 hours from LKN vs. >9 to ≤24 hours from LKN).

3.3 Study Drug Compliance and Time on Study

Subjects will receive one dose of study drug per the protocol. Number of subjects dosed and number who received the entire infusion of study drug will be presented under subject accountability.

Time on study, based on number of days from date of infusion to the last date on study will be summarized by treatment group and overall using summary statistics. If a subject is not dosed, the date of randomization will be used. The last date on study will be taken as the last visit/evaluation dates from all available data for the subject. If a subject dies during the study, the last date on study will be the date of death.

3.4 Concomitant Therapy

All concomitant medications will be coded using the World Health Organization (WHO) medication dictionary. Concomitant non-drug therapies will be coded using Medical Dictionary for Regulatory Activities (MedDRA).

A concomitant therapy (including medication or non-drug therapy) is defined as any therapy that is taken on or after the date/time of the randomized dose. This includes therapies that are started prior to the date/time of randomized dose if their use continues on or after Day 0. If a subject never receives the randomized dose, date of randomization will be used.

In order to define concomitant for therapies with missing start or stop date, the following additional criteria are defined:

- if both the start and stop dates of a particular therapy are missing, that therapy is considered concomitant;
- if the start date of a therapy is missing and the stop date of that therapy falls on or after the date of randomized dose, that therapy is considered concomitant;
- if the start date of a therapy is prior to the date/time of randomized dose and the stop date of that therapy is missing and the therapy is listed as continuing, that therapy is considered concomitant; or
- if the start date of a therapy is prior to the date/time of randomized dose and the stop date of that therapy is missing and the therapy is listed as *not* continuing, that therapy is considered *not* concomitant.

Concomitant medications will be summarized separately as the number and percentage of subjects taking any concomitant medications and the number and percentage of subjects taking each individual concomitant medication. Concomitant non-drug therapies will be summarized similarly.

3.5 Protocol Deviations

Protocol deviations identified during site monitoring will be captured in a protocol deviation log. These deviations will be listed. Also the number and percentage of subjects that meet the exclusion criteria from the per protocol population will be summarized.

4 Pharmacokinetic Data

Data collected from subjects in both treatment windows (≤9 or >9 and ≤24 hours from LKN) will be pooled in all analyses in Section 4 unless specified otherwise.

Analysis Population

The pharmacokinetics (PK) population will include subjects who have received natalizumab at the 0 Hours Visit and had at least 1 measurable sample collected for the determination of natalizumab concentrations. Subjects will be analyzed by the treatment they actually received. If the infusion does not complete, the actual dose administered should be recorded to include the patient in the PK population.

Methods of Analysis

Natalizumab concentrations from samples collected at each scheduled timepoint will be summarized by treatment group with summary statistics (n, mean, standard deviation, %

coefficient of variation [%CV], standard error, geometric mean, geometric %CV, and median and range) for those in the PK population. Patients who do not receive full dose will be excluded from summary statistics. Number and percentage of subjects with values of BLQ will also be provided. The mean concentrations (±standard error [SE]) will be plotted over time on both a linear and a logarithmic scale for each treatment group. Only mean and median will be summarized or plotted in a given arm with less than 3 subjects with data.

Data will be summarized by the scheduled visits.

In general, for data that are summarized by visit, data from early termination visit will be assigned to an appropriate scheduled visit by using a windowing scheme as follows.

If a subject withdraws after the infusion but prior to the first scheduled visit, data from the early termination visit will be assigned to the first scheduled visit. For all other visits, the lower bound and the upper bound for the visit windows are defined as the midpoints of the scheduled visits. If the date and time from early termination visit falls in between the lower bound and the upper bound for a scheduled visit, then it will be assigned to that visit. The start date and time of the infusion will be the reference point (0 hour).

If more than one observation is within the same window, data from the regularly scheduled visit will be used in the summary statistics and analyses. If more than 1 observation exists from a regularly scheduled visit, the earlier visit will be used in the summary statistics and analyses.

Natalizumab concentrations will be analyzed by noncompartmental methods. The following pharmacokinetic parameters will be summarized (n, mean, standard deviation, % coefficient of variation [%CV], standard error, geometric mean, geometric %CV, and median and range) :

- C_{max}
- Time to C_{max} (t_{max})
- Area under the serum concentration versus time curve from dosing (time=0) to 120 hours after dosing ($AUC_{0\text{-}120h}$)
- Area under the serum concentration versus time curve from dosing (time=0) to 672 hours after dosing ($AUC_{0\text{-}672h}$)
- Area under the serum concentration versus time curve from dosing (time=0) to infinity after dosing (AUC_{inf})
- Area under the serum concentration versus time curve, from dosing (time=0) to last measurable concentration ($AUC_{0\text{-}last}$)
- Half-life ($t_{½}$)
- Time of last measurable concentration (t_{last})
- Volume of distribution at steady state (Vd_{ss})
- Clearance (CL)

Natalizumab concentrations and PK parameters data will be listed.

5 Pharmacodynamic Data

Data collected from subjects in both treatment windows (≤9 or >9 and ≤24 hours from LKN) will be pooled in all analyses in Section 5 unless specified otherwise.

Analysis Population

The pharmacodynamics (PD) population will include subjects who have received the entire infusion of study treatment and have at least 1 post-baseline PD assessment. Subjects will be analyzed in the treatment group they actually received. If the infusion does not complete, the actual dose administered should be recorded to include the patient in the PD population.

Methods of Analysis

Whole blood samples collected during the study may be used to determine α4 integrin saturation and α4 integrin expression on leukocyte subsets, including but not limited to monocytes, lymphocytes, neutrophils, dendritic cells, T cells, and B cells using flow cytometry. The following will be summarized:

1. Maximum % saturation [R_{max}], time to R_{max} [TR_{max}]
2. Percent saturation pre-dose [$R_{pre\text{-}dose}$] and percent saturation at the end of the dosing interval [R_{672}] will be calculated.

Where applicable, summary statistics (n, mean, standard deviation, % coefficient of variation [%CV], standard error, geometric mean, geometric %CV, and median and range) and change from baseline will be presented by treatment group at each scheduled timepoint. Patients who do not receive full dose will be excluded from summary statistics. Mean values will be plotted over time for each PD marker. Summary statistics and plots will not be summarized in a given arm with less than 3 subjects with data. The mean % saturation values (±SE) will be plotted over time for both the entire study population and the individual doing groups. No formal statistical testing for treatment comparison will be performed.

Data will be summarized by the scheduled visits.

In general, for data that are summarized by visit, data from early termination visit will be assigned to an appropriate scheduled visit by using a windowing scheme as follows.

If a subject withdraws after the infusion but prior to the first scheduled visit, data from the early termination visit will be assigned to the first scheduled visit. For all other visits, the lower bound and the upper bound for the visit windows are defined as the midpoints of the scheduled visits. If the date and time from early termination visit falls in between the lower bound and the upper bound for a scheduled visit, then it will be assigned to that visit. The start date and time of the infusion will be the reference point (0 hour).

If more than one observation is within the same window, data from the regularly scheduled visit will be used in the summary statistics and analyses. If more than 1 observation exists from a regularly scheduled visit, the earlier visit will be used in the summary statistics and analyses.

6 Efficacy Data

Analysis Population

The intent-to-treat (ITT) population will include subjects who are randomized.

The modified intent-to-treat (MITT) population will include subjects who are randomized and have received the entire infusion of study treatment. Subjects who were accidentally enrolled based on conditions that mimicked stroke symptom at presentation will be excluded from the MITT population. Subjects will be analyzed in the group to which they are randomized.

The per-protocol population (PP) will include subjects from the MITT population without major inclusion/exclusion criteria violations (specified below). The criteria for exclusion from the PP population are listed below:

1. Clinical diagnosis of supratentorial acute ischemic stroke is >24 hours from LKN prior to study treatment initiation.
2. Score of NIHSS at screening is <5 or > 23 for subjects initiating treatment ≤9 hours from LKN.
3. Score of NIHSS at screening is <5 or > 15 for subjects initiating treatment >9 to ≤24 hours from LKN.

4. Severe stroke defined by imaging criteria based on either one of the following
    a. ASPECTS score of 0 to 4 based on head Computed Tomography (CT) OR
    b. Acute infarct volume on Magnetic Resonance Imaging (MRI) diffusion weighted imaging ≥70 mL (cc).

In addition, if there are subjects who are unblinded unexpectedly during the study, the efficacy data for these subjects may be censored after the date of unblinding.

6.1 General Methods of Analysis

Summary statistics will be presented. For continuous endpoints, the summary statistics will generally include: number of subjects with data, mean, standard deviation, median, quartiles, and range. For categorical endpoints, the summary statistics will generally include: number of subjects in corresponding analysis population, number and percentage of subjects in each category.

In general, for data that are summarized by visit, data from the early termination visit will be assigned to an appropriate scheduled visit by using a windowing scheme as described below.

If a subject withdraws after the infusion but prior to the first scheduled visit, data from the early termination visit will be assigned to the first scheduled visit. For all other visits, if the date and time from early termination visit falls in between the lower bound and the upper bound for a scheduled visit, then it will be assigned to that visit. The lower bound and the upper bound for scheduled visit windows are defined as the midpoints of two scheduled visits. The start date and time of the infusion will be the reference point (0 hour).

If more than one observation is within the same window, data from the regularly scheduled visit will be used in the summary statistics and analyses. If more than 1 observation exists from a regularly scheduled visit, the earlier measurement will be used in the summary statistics and analyses.

If more than 10% of data are missing for any baseline covariate, that covariate will not be used in the modeling described below.

Treatment included in statistical modeling will be used as a categorical variable with three levels (300mg, 600mg and placebo), if not specified otherwise.

All efficacy endpoints will be evaluated on the MITT population. The analyses performed on the MITT population will be considered the primary analyses.

Primary and selected secondary efficacy endpoints (mRS, BI) will also be analyzed based on the per-protocol population.

The final model for each analysis in the MITT population will also be used for the PP analysis and any subgroup analyses.

Subjects will be analyzed in the group to which they are randomized for the MITT analysis and based on actual treatment received in the PP analysis.

If the number of subjects in a certain subgroup is too small (e.g., < 10% of total number of subjects in the MITT population), the analysis for that subgroup may not be performed.

Statistical analyses on the primary and secondary endpoints are summarized in Appendix A.

6.2 Primary Endpoints

Composite global measure of functional disability based on a score of 0 or 1 on the modified Rankin scale (mRS) and a score of ≥95 on the Barthel Index (BI) at Day 90.

It should be noted that this is not based on a single composite score by summing score for each measure. Instead, each measure will result in a 0 or 1 score which can be assessed by logistic regression in a univariate fashion. However, the primary estimate of odds ratio for overall improvement across mRS and BI will be derived based on a global model (see below). The same model and its associated Generalized Estimation Equation (GEE) method were used in reporting results from a global outcome analysis of neurologic and disability scores in the European Cooperative Acute Stroke Study (ECASS) III [1].

6.2.1 Primary Analysis

The global composite of an excellent outcome on the mRS (score of 0 or 1) and an excellent outcome on the BI (score of at least 95) at Day 90 will be analyzed using GEE models with the logit link function. The two components will be the within-subject repeated measures in the GEE model. Treatment will be included in statistical modeling as a categorical variable with three levels (300mg, 600mg and placebo). The covariates to be included in the models are:

- Baseline NIHSS category (score 5-15, 16-23),
- tPA use (yes/no),
- Thrombectomy procedure (yes/no),
- Age (<60, 60-69, 70-80),
- Treatment window (≤9 vs. >9 and ≤24 hours from LKN)
- Region (UK/Germany vs. Spain vs. US)

Unstructured working correlation structure between the two components will be used in the GEE models. Missing Day 90 mRS and BI will be imputed using multiple imputation methodologies (with 50 replications). Further details of the imputation procedure are provided in Section 6.2.2.

Step 1: Assessment of Interaction Between Treatment and Treatment Window in Analysis of Pooled Data from Both Treatment Windows (≤9 vs. >9 and ≤24 hours from LKN)

The first model will include the covariates described above and the two-way interaction of treatment and treatment window.

Step 2

If the p-value of the interaction term between treatment and treatment window is ≥ 0.3 for the combined active dose groups versus placebo, the primary analysis population will include subjects in both treatment windows, and a main effect only GEE model with the same covariates as described above will be the basis of the primary analysis.

If the p-value of the interaction is <0.3, the primary analysis population will include subjects in the ≤9 hours from LKN window only. A main effect only GEE model with the same covariates except treatment window will be the basis of the primary inference. A separate analysis for subjects with treatment window >9 and ≤24 hours from LKN will also be performed.

Primary Inference

Consistent with the protocol, depending on the primary treatment window based on the interaction test, the global odds ratio in the primary analysis model, either based on pooled treatment windows or ≤ 9 hours from LKN, the composite odds ratio of achieving excellent outcomes in mRS and BI of active (two doses combined) versus placebo will be derived from the GEE model with 95% confidence interval. A conventional P-value will be provided for descriptive purpose but will not be the basis for inference.

Characterization of the primary comparison of combined active dose groups vs. placebo for the primary endpoint will be made utilizing a Bayesian framework. Posterior probability of treatment effect in terms of odds ratio of active (combined doses) versus placebo > 1.2 will be derived based on observed results in this study 101SK202 (ACTION 2) with the prior information from Study 101SK201 (ACTION) incorporated as prior distribution. As the functional endpoints mRS and Barthel Index were not primary endpoints of study 101SK201 (ACTION) and the composite of mRS and BI was one of the four related functional endpoints, the findings in ACTION are likely subject to uncertainty beyond the estimated distribution nominally suggested. To account for this additional uncertainty, the prior distribution derived from study 101SK201 (ACTION) will be down-weighted by assuming the sample size of ACTION was 50% of the actual and thereby increasing the variance of the prior distribution. Additionally, an "initial prior" probability distribution describing the general likelihood of success in Phase 2 acute ischemic stroke trials based on literature will also be applied. A probability distribution that gives 30% chance for the odds ratio> 1.2 and 14% chance for the odds ratio > 1.45 (assuming a normal distribution of log odds ratio) is utilized based on a study of clinical development success rates 2006-2015 [2].

The resulting prior probabilities of odds ratio > 1.2 before the study 101SK201 (ACTION) and subsequently before the current study 101SK202 (ACTION 2) are summarized in Appendix B.

Additionally, pooled analysis of the results of the 101SK202 study and 101SK201 study will be performed to create a summary estimate of the treatment effect.

Other comparisons based on the primary model
The same primary GEE model will be used to compare each dose level to placebo and between dose levels with the following rank ordering:

1. 600 mg versus placebo
2. 300 mg versus placebo
3. 600 mg versus 300 mg

6.2.2 Handling of Missing Values

6.2.2.1 Multiple Imputation for missing mRS

Assessment of mRS is made on Day 5, 30, and 90. Death is scored as 6 in mRS and thus there will be no missing mRS values due to death. For missing mRS values due to other reasons, an assumption of missing at random will be checked and, if considered valid, the missing values will be imputed based on a linear mixed model performed on the mRS scores leveraging the information of the patient at other time points as well as the time patterns of all other patients. The model will include treatment as a classification variable and the covariates described above in Section 6.2.1. Multiple imputations will be performed using PROC MI in SAS. Fully conditional specification method will be used in this SAS procedure. Imputed values will be rounded to the closest valid mRS score. A total of 50 imputations will be conducted with random seed pre-specified; each results in a complete dataset for all subjects. The point and interval

estimates of the odds ratio from the primary analysis models will be obtained using PROC MIANALYZE.

6.2.2.2 Multiple Imputation for missing BI

For missing value due to death, BI will be imputed by the worst score, 0. For other missing values, the same method as described in Section 6.2.2.1will be applied to BI.

6.2.3 Sensitivity Analysis of the Primary Endpoint

Sensitivity analyses will be conducted on the same analysis population with respect to treatment window as for the primary analysis.

6.2.3.1 Sensitivity Analysis of composite global measure of functional disability based on a score of 0, 1 or 2 on mRS and a score of ≥85 on BI at Day 90

The same analysis in Section 6.2.1 using the same method for missing value handling will be conducted to assess the treatment effect on good outcomes defined as Day 90 mRS score of 0,1, or 2 and BI score ≥85.

6.2.3.2 Sensitivity Analysis using last-observation-carry-forward (LOCF) procedure

<u>mRS</u>

If a subject is known to be dead on or before a specific post-baseline visit, the missing mRS at that specific visit and all remaining visits will still be set to the worst possible outcome. Otherwise, the last post-baseline mRS (LOCF) will be used.

<u>BI</u>

If a subject is known to be dead on or before a specific post-baseline visit, the missing BI at that specific visit and all remaining visit will still be set to the worst possible outcome. Otherwise, the last post-baseline BI (LOCF) will be used.

6.2.3.3 Sensitivity Analysis of data as observed

Missing values of mRS and BI as a result of death will be imputed by the worst respective scores as discussed in Section 6.2.2. No other imputation of missing values will be conducted.

6.2.3.4 Sensitivity Analysis Using Per-Protocol (PP) Population

The primary analysis as described in Section 6.2.1 with the same method to handle missing values as in Section 6.2.2 will be performed on the PP population.

6.2.3.5 Sensitivity Analysis incorporating Day 90 NIHSS score

A composite global measure based on a score of 0 or 1 on mRS, a score of ≥95 on BI and a score of 0 or 1 on NIHSS at Day 90 will be analyzed. Similar analyses as in Section 6.2.1 will be conducted. Missing values for NIHSS will also be imputed using multiple imputation as described in Section 6.2.2. If a subject is known to be dead due to stroke on or before a specific post-baseline visit, missing NIHSS scores at that specific visit and all remaining visits will be imputed to the worst possible score (42).

6.2.4 Supportive analysis of primary endpoint on Day 5 and 30

Composite global measure of functional disability based on a score of 0 or 1 on the modified Rankin scale (mRS) and a score of ≥95 on the Barthel Index (BI) at Day 5 and Day 30 will be analyzed separately. The same statistical analysis as described in Section 6.2.1 and the same method to handle missing values described in Section 6.2.2 will be performed on the MITT population.

6.2.5 Subgroup Analysis

Subgroup analyses will be performed for the Day 90 primary endpoint following the same imputation and statistical analysis as described in Sections 6.2.1 and 6.2.2 for the primary analysis. These analyses will be performed on the MITT population only. Forest plots will be used to present results across subgroups. The following subgroups will be evaluated in the analysis of primary endpoint:

- tPA use prior to infusion of study drug (yes/no)
- Thrombectomy procedure (yes/no)
- Baseline NIHSS categories (NIHSS score 5-15, 16-23) for the ≤9 hours from LKN window only (enrollment in >9 and ≤24 hours from LKN window is confined to patients with baseline NIHSS score 5 to 15).
- Treatment time window (≤ 9 hours or > 9 to ≤ 24 hours from the subjects' LKN)
- Country
- Age (<60, 60-69, 70-80)
- sex

If the number of subjects in a certain subgroup is too small (e.g., < 10% of total number of subjects in the MITT population), the analysis in that subgroup may not be performed.

6.2.6 Dose-Response and Exposure-Response

The relationship between the excellent outcome on mRS and BI and dose level will be evaluated. The Cochran-Armitage trend test of a monotonically increasing dose response in proportion of excellent outcome on mRS and BI will be performed. Dose level will be log-transformed. For placebo treatment, the dose level will set to small value (1) after the log transformation.

Exposure-response (ER) models, e.g., sigmoid Emax model and its special case Emax model where hill parameter is 1, will be explored to quantify the effects of exposure (using AUC from dosing to last measurable concentration) on response variables (mRS and BI at Day 90). Covariates of age, body weight, baseline NIHSS category (score 5-15, 16-23), tPA use (yes/no), thrombectomy (yes/no), and treatment window may be explored in the nonlinear ER models.

6.2.7 Other analyses

Other analyses may be conducted to explore factors that might potentially influence patients' outcome or impact the treatment effect, e.g., history of diabetes and baseline glucose level.

6.3 Secondary Endpoints

The five secondary endpoints will be assessed as following:

- mRS at Day 90 (excellent outcome, good outcome and shift analysis)
- BI score at Day 90 (excellent outcome, good outcome and full scale)
- Montreal Cognitive Assessment (MOCA) at Day 90 (categorical and continuous analysis)
- Stroke Impact Scale-16 (SIS-16) at Day 90 (categorical and continuous analysis)
- NIHSS score at Day 90 (categorical and continuous analysis)

Characterization of each secondary endpoint using the methods described for the primary endpoint will be based on the same analysis population with respect to treatment window as for the primary endpoint. The same covariates as in the primary analysis for the primary endpoint will be included.

6.3.1 mRS at Day 90

Summary statistics of the mRS will be presented by time point, treatment group and treatment window for observed data. In addition, the number and percentage of each score will be presented by time point, treatment group and treatment window.

6.3.1.1 Analysis of excellent outcome and good outcome

Same analysis approach described for the primary endpoint including missing value handling, sensitivity analyses, supportive analyses, and subgroup analyses described in Section 6.2 will be applied to the excellent outcome in mRS.

Similar primary analysis as described in Section 6.2 will be performed on the proportion of subjects who have a good outcome (0, 1 or 2) on Day 90, including missing value handling. In addition, an exploratory analysis will be performed to further understand the treatment effect on unfavorable outcome (mRS 0-4 versus 5-6)

6.3.1.2 Analysis of mRS shift

Ordinal logistic regression

Same analysis approach described for the primary endpoint including primary inference framework, missing value handling, sensitivity analyses, supportive analyses, and subgroup analyses described in Section 6.2 will be applied to the shift analysis in mRS using ordinal logistic regression models. In addition to covariates described in Section 6.2, history of diabetes as a covariate may be explored as well.

Van Elteren's test

A non-parametric test on the mRS shift will also be performed using Van Elteren's test. Grotta bar charts for the distribution of mRS will be presented.

The mRS distribution at Day 30 will also be analyzed similarly.

6.3.1.3 Repeated Measure Modeling Based on all visits

A mixed model for repeated measure (MMRM) of ordinal mRS data that includes all time points (Day 5, Day 30 and Day 90) will be conducted. Comparison between each active dose level vs. placebo will be performed on all visits. Missing values as a result of death will be handled in the same way as described in Section 6.2.2. Treatment by visit interaction will be included in the model as explanatory variables. The same covariates as in the primary analysis for the primary endpoint will also be included. Random intercept for subject will be used in the model. Pairwise comparisons between dose levels and two doses combined by equal weights vs. placebo will be performed on all visits.

6.3.2 BI score at Day 90

Summary statistics of the Barthel Index will be presented by time point, treatment group and treatment window for observed data.

6.3.2.1 Analysis of excellent outcome

Same analysis approach described for the primary endpoint primary inference framework, missing value handling, sensitivity analyses, supportive analyses, and subgroup analyses described in Section 6.2 will be applied to the excellent outcome in BI.

6.3.2.2 Analysis of good outcome (BI ≥85)

Similar primary analysis as described in Section 6.2 will be performed on the proportion of subjects who have a good outcome (BI ≥85) on Day 90, including missing value handling. The sensitivity analysis using LOCF will be conducted in the same way as described in Section 6.2.3.2.

6.3.2.3 Repeated Measure Analysis

A MMRM analysis that includes all time points (Day 5, Day 30 and Day 90) will be conducted. Pairwise comparisons between dose levels and two doses combined by equal weights vs. placebo will be carried out on all visits. Missing values as a result of death will be handled in the same way as described in Section 6.2.2. Treatment by visit interaction will be included in the model as explanatory variables. The same covariates as in the primary analysis for the primary endpoint will also be included. An unstructured variance-covariance matrix will be used in the model. In addition, a line plot of adjusted mean score at Day 5, 30 and 90 will be presented by treatment group.

6.3.3 Score on Montreal Cognitive Assessment (MOCA) at Day 90

The total raw score on the MoCA is calculated from the simple sum of scores of the 7 components. If a subject has 12 years of education or fewer, a point is added to the total score, except if a subject scores 30/30, in which case a point is not added.

Summary statistics of MoCA at each visit will be presented by treatment group and treatment window for observed data.

In addition, MoCA assessments will be presented using summary statistics for the following categories: < 10 (severe cognitive impairment), 10-17 (moderate cognitive impairment) and >= 18 (mild cognitive impairment).

6.3.3.1 Analysis on proportion of subjects with MoCA score ≥ 22, 24 and 26 at Day 90

The proportion of subjects who have a MoCA score ≥ 22, 24 and 26 at Day 90 will also be evaluated separately. Similar logistic regression analysis described for the individual component of the primary endpoint including primary inference framework, missing value handling for death described in Section 6.2 will be applied. The only exception is that single imputation rather than multiple imputation will be used.

Missing Value Handling

A single imputation procedure will be performed using data collected from subjects. A MMRM analysis of observed data on all visits will be used to impute missing on Day 90. Data due to death will be assigned as missing in this single imputation. The model will include the same

covariates as for the primary analysis for the primary endpoint and treatment by visit interaction. An unstructured variance-covariance matrix will be used in the model.

Supportive analysis will be performed on MoCA collected on Day 5 and 30, following the same analysis method as for Day 90.

6.3.3.2 Repeated Measure Analysis of MoCA Scores Collected on All Visits

The same repeated measure analysis of MoCA will be conducted as described in Section 6.3.2.3.

6.3.4 Stroke Impact Scale-16 (SIS-16) at Day 90

A patient's score on the SIS-16 is calculated using a multi-step algorithm. First scores for non-missing items are summed. This sum is then standardized by subtracting the lowest possible score (1) and dividing by the range of possible scores (4), according to the following equation:

Score = [(sum score – minimum possible score)/ (maximum possible score – minimum possible score)]*100

Scores are only evaluated if the number of non-missing items is higher than 50% (i.e >= 9). If 8 or more items are missing, a valid SIS-16 total score will not be computed.

Summary statistics of SIS-16 at each visit will be presented by treatment group and treatment window for observed data.

6.3.4.1 Analysis on proportion of subjects with SIS-16 ≥ median at Day 90

The same set of analyses will be conducted on the proportion of subjects who have SIS-16 score ≥ median of all subjects at Day 90 as described in Section 6.3.3.1.

6.3.4.2 Repeated Measure Analysis of SIS-16 Collected on All Visits

The same set of repeated measure analyses of SIS-16 will be conducted as described in Section 6.3.3.2.

6.3.5 Change in NIHSS Score from Baseline to Day 90

Summary statistics of the NIHSS Score and number with excellent outcome according to criterion 1 (NIHSS of 0 or 1) and criterion 2 (NIHSS of 0 or 1, or at least an 8-point improvement from Baseline) will be presented by time point, treatment group and treatment window using observed data.

6.3.5.1 Analysis of Excellent Outcome on Day 90

Same analysis approach described for the primary endpoint including primary inference framework, missing value handling, sensitivity analyses, supportive analyses, and subgroup analyses described in Section 6.2 will be applied to each of the two definitions of excellent outcome in NIHSS. Missing values known as a result of death due to stroke will be handled in the same way as described in Section 6.2.3.5.

6.3.5.2 Repeated Measure Analysis of NIHSS Collected on All Visits

Change in the NIHSS score will be modeled by a repeated measures mixed model jointly for 24 Hours, Day 5, 30 and 90. Treatment by visit interaction will be considered as a covariate in the model. The same covariates as in the primary analysis for the primary endpoint will also be included. Treatment will be included as a categorical variable with three levels (300mg, 600mg and placebo). Comparison between each active dose level vs. placebo and two doses combined by equal weights vs. placebo will be performed on all visits. Missing values known as a result of death due to stroke will be handled in the same way as described in Section 6.2.3.5. An unstructured variance-covariance matrix will be used in the model. Treatment contrasts at 24 Hours, Day 5, 30 and 90 will be obtained along with their 95% confidence intervals.

6.4 Exploratory Endpoints

General Consideration for Exploratory Endpoints

If a subject is known to be dead on or before a specific post-baseline visit, the missing at that specific visit and all remaining visit(s) will be set to the worst possible outcome before conducting any MMRM analysis in Section 6.4.

Analysis on each exploratory endpoint will be based on the same analysis population with respect to treatment window as for the primary endpoint.

Functional Independence Measure score

Functional Independence Measure (FIM) is a widely used functional performance measure developed specially for the inpatient acute rehabilitation population. FIM is an 18-item instrument graded on a 7-point ordinal scale (1= Total Assistance (subject = less than 25%), 7= Complete Independence (timely, safely)), with a maximum total score of 126. The 7-point ordinal scale indicates the burden of care associated with each aspect of function.

Summary statistics of total score FIM at each visit will be presented by treatment group and treatment window for observed data. In addition, the descriptive summary for subtotal score for each function will be represented by treatment and treatment window as well.

Symbol Digit Modalities Test score –Total correct

Symbol Digit Modalities Test (SDMT) measures informational processing speed in the visual modality. Patients are presented with a key that includes 9 numbers and each paired with a differed symbol. The patients will provide the correct numbers that accompany the symbols.

Summary statistics of SDMT for the total number of responses, total correct number and the proportion of correct responses at each visit will be presented by treatment group and treatment window using observed data.

A MMRM analysis on observed data that includes all time points (Day 5, Day 30 and Day 90) will be used. Treatment by visit interaction will be included in the model as explanatory variables. The same covariates as in the primary analysis for the primary endpoint will also be included. An unstructured variance-covariance matrix will be used in the model. Comparison between each active dose level vs. placebo will be performed on all visits. At each time point, adjusted mean for each treatment arm and treatment contrasts will be derived along with their 95% confidence intervals.

Fatigue Severity Scale score

The Fatigue Severity Scale (FSS) is a 9-item scale which measures the severity of fatigue and its effect on a person's activities and lifestyle in patients with a variety of disorders. The items are scored on a 7 point scale with 1 = strongly disagree and 7= strongly agree. The minimum total score is 9, and the maximum score possible is 63, with the higher the score indicating greater fatigue severity.

The descriptive statistics of total score at each visit will be summarized by treatment group and treatment window using observed data.

A MMRM analysis on observed data will be conducted as for SDMT.

Beck Depression Inventory 2 score

The Beck Depression Inventory 2 (BDI-2) is a self-report inventory and can be self-administered or verbally administered. The instrument rates items on a 4 point scale that ranges from 0-3. Ratings are summed to provide a total score rating from 0 to 63.

Summary statistics of BDI-2 at each visit will be presented by treatment group and treatment window for observed data.

In addition, the number and percentage of subjects will be presented for the following categories: Non (0-13), Mild (14-19), Moderate (20-28) and severe (29-63).

A MMRM analysis on observed data will be conducted as for SDMT.

Subject direct resource use (assessed using a health resource utilization questionnaire)

The Health Resource Utilization (HRU) questionnaire quantifies the amount of time patients spent in various settings of care, number of times to visit professional healthcare, the readmission rate and rehabilitation utilization.

Descriptive summary statistics for health resource utilization at each visit that were collected at each visit will be presented by treatment group and treatment window using observed data.

EuroQoL EQ-5D-3L (questionnaire)

The Euroqol (EQ)-5D-3L is a generic health-related quality of life (QoL) instrument which has been extensively validated. A Health State Profile is used in this protocol. With the Health State Profile, patients record their level of current health for five dimensions comprising a health profile: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. A scoring formula developed by the EuroQol Group is then used to assign utility values for each patient's Health State Profile. A score of 1, 2, or 3 are possible responses for each of the five questions in the Health State Profile (1=no problems, 2=some problem, 3=severe problems). The digits for the five dimensions are combined in a 5-digit number describing a person's health state. A summary index will be derived from the 5 questions by conversion with a table of scores. A subject assessment of health scale is also recorded.

Summary statistics for the EQ-5D for the converted index score and subject assessment of health scale at each visit will be presented for observed data. Also the number and percentage of patients of possible response in each of the five questions will be summarized by treatment group and treatment window.

A MMRM analysis on converted scores from observed data will be conducted as for SDMT.

The same MMRM analysis will be performed on subject assessment of health scale.

7 Safety Data

Analysis Population

Data collected from subjects in both treatment windows (≤9 or >9 and ≤24 hours from LKN will be pooled in all analyses in Section 7 unless specified otherwise.

The safety population is defined as subjects who have received any study treatment, including both cases of complete and incomplete infusions. Subjects will be analyzed in the treatment group for the treatment they actually received, regardless of whether the subject completed the infusion. All safety analyses will be based on the safety population.

7.1 Adverse Events

Methods of Analysis

All AEs will be collected according to the protocol. These AEs will be classified using the Medical Dictionary for Regulatory Activities (MedDRA: the latest version will be used at the time of analysis). A treatment-emergent AE is defined as any AE that has onset on or after the

first dose of study treatment, or any pre-existing condition that has worsened after the first dose of study treatment.

All clinical adverse events will be analyzed based on the principle of treatment emergence.

In order to define treatment emergence for events with missing start or stop dates the following additional criteria will be used:

- if both the start and stop dates for a particular event are missing, then that event is considered treatment-emergent;
- if the start date for a particular event is missing and the stop date/time fall after the start date/time of the study dose, then that event is considered treatment-emergent;
- if the start date was the same as the dose date and the start time was missing, and the stop date/time is after the date/time of dose or cannot be compared with the date/time of dose, then that event is considered treatment-emergent.

For events with a partial start date, the year/month of the event date will be compared to that of the dosing date to determine whether the event is treatment-emergent.

The incidence of treatment emergent adverse events will be summarized for any events, events of various severity, events with relationship to study treatment, serious events, and events leading to study drug discontinuation or study withdrawal.

Treatment-emergent AEs will be summarized separately by treatment group as follows:

- by preferred term
- by primary system organ class
- by primary system organ class and preferred term
- by severity, primary system organ class and preferred term
- by relationship to study treatment, primary system organ class and preferred term
- by primary system organ class and preferred term for serious adverse events

A listing of the following will be presented:

- serious adverse events
- AEs leading to discontinuation of study drug
- AEs leading to withdrawal from study
- Deaths

The incidence of treatment-emergent adverse events occurring in at least 5% in any treatment group will be presented by preferred term.

For the analysis of incidence by severity, the occurrence of the AE with the greatest severity will be used, and a subject will be counted only once and only in the category of the greatest severity

for each event. For the analysis of incidence by relationship to study treatment, the occurrence of the AE with the strongest relationship to study treatment will be used and a subject is counted only once and only in the category of the strongest relationship to study treatment for each event.

Recurrent stroke will be summarized by treatment group. Relevant information including timing of the recurrent stroke relative to study treatment will be provided.

Any deaths that occurred during the study will be listed. Relevant information including timing of the death relative to study treatment and the investigator assessment of the cause of death will be provided. The incidence of death occurring on or prior to Day 5, Day 30 and Day 90 will be presented by treatment group.

In addition, subgroup analysis will be performed for selected safety endpoints including SAEs based on the following subgroups:

- treatment time window (≤ 9 hours or > 9 to ≤ 24 hours from the subject's LKN)
- tPA use prior to infusion of study drug (yes versus no)
- Baseline NIHSS categories
- Age (<60, 60-69, 70-80)
- sex

If the number of subjects in a certain subgroup is too small (e.g., < 16 subjects), the analysis in that subgroup may not be performed.

In addition, adverse events of special interest will be identified by pre-specified list of coded terms. For each AE of special interest, the number and percentage of subjects with event will be summarized by preferred term.

7.2 Laboratory Data

The main analyses of laboratory data will focus on analyses of data from baseline to post-baseline. Baseline is defined as the closest non-missing value prior to the infusion of study treatment.

Laboratory data will be summarized using shift tables where appropriate. Each subject's hematology and blood chemistry values will be flagged as "low", "normal", or "high" relative to the normal ranges of the central laboratory or as "unknown" if no result is available.

Shifts from baseline to high/low status for hematology and blood chemistry parameters will be presented. In each summary, the denominator for the percentage is the number of patients at risk for the shift. The number at risk for shift to low is the number of subjects whose baseline value was not low and who had at least one post-baseline value. The number at risk for shift to high is the number of subjects whose baseline value was not high and who had at least one post-baseline value. Subjects will be counted only once for each parameter and each type of shift regardless of

how many post-dosing assessments had that type of shift. Subjects with shift to low or high will be listed by laboratory parameter and shift type.

Summary statistics for actual values and changes from baseline will also be summarized by treatment group and overall by timepoint.

In addition, the shift from baseline to the maximum post-baseline value and the shift from baseline to the minimum post-baseline value will be presented for each laboratory test by treatment group. The rationale for using the minimum/maximum values or worst values is that should a treatment affect a laboratory value, that value could be affected at different times for different subjects. Therefore, these analyses present the most extreme values for each subject over time.

Analysis of liver Function Tests

A summary of the number and percentages of subjects meeting the laboratory abnormality criteria listed below will be provided:

- ALT > 3xULN
- ALT > 5xULN
- AST > 3xULN
- AST > 5xULN
- AST or ALT > 3xULN
- AST or ALT > 5xULN
- Total Bilirubin > 2xULN
- ALP >1.5xULN
- AST or ALT > 3x ULN and Total Bilirubin > 2x ULN

7.3 Vital Sign Data

The analysis of vital signs will focus on the incidence of clinically relevant abnormalities.

The number of subjects evaluated and the number and percentage of subjects with clinically relevant post-baseline abnormalities will be presented by treatment group. The criteria for clinically relevant post-baseline abnormalities are shown in the following table. Summary statistics for actual values and change from baseline will also be presented.

Table 1 Criteria to Determine Clinically Relevant Abnormalities in Vital Signs

Vital Sign	Criteria for Abnormalities
Temperature	> 38°C and an increase from pre-dose of at least 1°C
Heart rate	> 120 beat per minute and an increase from pre-dose of more than 20 beats per minute. < 50 beats per minute and a decrease from pre-dose of more than 20 beats per minute.
Systolic Blood Pressure	> 180 mmHg and an increase from pre-dose of more than 40mmHg. < 90 mmHg and a decrease from pre-dose of more than 30 mmHg.
Diastolic Blood Pressure	> 105 mmHg and an increase from pre-dose of more than 30 mmHg. < 50 mmHg and a decrease from pre-dose of more than 20 mmHg.
Respiratory rate	>=20 breaths per minute with a pre-dose rate of < 20 breaths per minute <= 10 breaths per minute with a pre-dose rate of > 10 breaths per minute

8 Immunogenicity Data

Analysis Population

The immunogenicity population will include subjects who have received any portion of the infusion of study treatment and have at least 1 sample available for immunogenicity analysis. Subjects will be analyzed in the treatment group they actually received.

Methods of Analysis

Anti-natalizumab antibodies are tested at Day 30 and at Day 90 (or at Early Termination Safety Follow-up Visit).

The number and percentage of subjects with a positive anti-natalizumab antibody result will be presented at each scheduled timepoint by treatment group. The number and percentage of subjects with at least one post-baseline positive anti-natalizumab antibody results will be summarized by treatment group. The number and percentage of subjects who have positive anti-natalizumab antibody results at two post-baseline visits will be summarized by treatment group.

9 Interim Analysis

An interim futility analysis may be performed after 50% of the study population has completed the Day 30 assessment. No interim stopping rules for superiority will be applied.

10 Sample Size Justification

A sample size of 270 (90 per treatment group) will provide at least 88% probability for the point estimate of the OR for the primary comparison of natalizumab (dose groups combined) versus placebo on the global composite measure at Day 90 to exceed 1.3 in both treatment windows (i.e., same efficacy for subjects being treated ≤9 hours and >9 to ≤24 hours from LKN), assuming a true OR of 1.8 as observed in Study 101SK201. Adopting a conservative assumption of a 50% reduction in treatment efficacy in the >9 to ≤24 hour treatment window when compared to the ≤9 hour treatment window, this probability will be at least 80%. An OR of ≥1.3 on the global outcome measure is considered to be clinically meaningful based on the effect of tPA in the 3- to 4.5-hour time window. At this sample size, the probability of observing a point estimate exceeding 1.3, if the true OR is ≤1 (i.e., the false positive rate), is less than 0.2. In addition, if the true OR of the higher dose versus placebo is 3.0, as observed in the subgroup with exposure above the median in Study 101SK201, the probability of observing an OR of ≥1.3 when comparing the 600- to 300-mg dose is at least 74%.

Appendix A: Summary of statistical analyses on the primary and secondary endpoints.

By default, for each analysis the overall treatment effect (300 mg and 600 mg combined) versus placebo as the primary comparison and all pairwise comparisons of the three dose levels will be conducted.

By default, all analyses on data pooled from both treatment windows will be conducted with and without the treatment by treatment window interaction.

Endpoint	Analysis type	Model	Analysis populati-on	Visit(s)	Treatment windows combined Y or N?	Missing Data Handling	Section Reference
Primary: Joint odds of mRS ≤ 1 and Barthel Index ≥ 95 at Day 90	Primary	GEE	MITT	Day 90	Y	MI of individual components except for death (worst score assigned)	6.2.1
	Sensitivity Different cutoffs of mRS and BI	GEE	MITT	Day 90	Y	Same as the primary.	6.2.3.1
	Sensitivity Different cutoffs of	GEE	MITT	Day 90	N	Same as the primary.	6.2.3.1

	mRS and BI						
	Sensitivity	GEE	MITT	Day 90	Y	LOCF for both components except for death (worst score assigned)	6.2.3.2
	Sensitivity	GEE	MITT	Day 90	N	LOCF for both components except for death (worst score assigned)	6.2.3.2
	Sensitivity	GEE	MITT	Day 90	Y	As Observed except for death (worst score assigned)	6.2.3.3
	Sensitivity	GEE	MITT	Day 90	N	As Observed except for death (worst score assigned)	6.2.3.3
	Sensitivity	GEE	PP	Day 90	Y	Same as the primary analysis	6.2.3.4
	Sensitivity NIHSS<=1 added as additional component	GEE	MITT	Day 90	Y	Same as the primary analysis	6.2.3.5
	Secondary	GEE	MITT	Day 90	N	Same as the primary analysis	6.2.1
	Supportive	GEE	MITT	Day 5/30 separately	N	Same as the primary analysis	6.2.4
	Subgroup (demographic and baseline)	GEE	MITT	Day 90	N	Same as the primary analysis	6.2.5
Secondary: Proportion of mRS at Day 90 ≤ 1	Primary	Logistic regression	MITT	Day 90	Y	MI of mRS ordinal result except for death (worst score assigned)	6.3.1.1
	Sensitivity	Logistic regression	MITT	Day 90	Y	LOCF except for death (worst score assigned)	6.3.1.1
	Sensitivity	Logistic regression	MITT	Day 90	Y	As Observed except for death (worst score assigned)	6.3.1.1
	Sensitivity	Logistic regression	PP	Day 90	Y	Same as the primary analysis	6.3.1.1
	Secondary	Logistic regression	MITT	Day 90	N	Same as the primary analysis	6.3.1.1
	Supportive	Logistic regression	MITT	Day 5/30 separately	N	Same as the primary analysis	6.3.1.1
	Subgroup (demographic and baseline)	Logistic regression	MITT	Day 90	N	Same as the primary analysis	6.3.1.1

Secondary: mRS shift	Primary	Proportional odds logistic regression	MITT	Day 90	Y	MI of mRS ordinal result except for death (worst score assigned)	6.3.1.2
	Sensitivity	Proportional odds logistic regression	MITT	Day 90	Y	LOCF except for death (worst score assigned)	6.3.1.2
	Sensitivity	Proportional odds logistic regression	MITT	Day 90	Y	As Observed except for death (worst score assigned)	6.3.1.2
	Sensitivity	Proportional odds logistic regression	PP	Day 90	Y	Same as the primary analysis	6.3.1.2
	Secondary	Proportional odds logistic regression	MITT	Day 90	N	Same as the primary analysis	6.3.1.2
	Secondary	Van Elteran's test	MITT	Day 90	Y	death assigned to worst score	6.3.1.2
	Secondary	Van Elteran's test	MITT	Day 90	N	death assigned to worst score	6.3.1.2
	Subgroup (demographic and baseline)	Proportional odds logistic regression	MITT	Day 90	N	Same as the primary analysis	6.3.1.2
Secondary: Proportion of BI at Day 90 ≥95	Primary	Logistic regression	MITT	Day 90	Y	MI of BI score except for death (worst score assigned)	6.3.2.1
	Sensitivity	Logistic regression	MITT	Day 90	Y	LOCF except for death (worst score assigned)	6.3.2.1
	Sensitivity	Logistic regression	MITT	Day 90	Y	As Observed except for death (worst score assigned)	6.3.2.1
	Sensitivity	Logistic regression	PP	Day 90	Y	Same as the primary analysis	6.3.2.1
	Secondary	Logistic regression	MITT	Day 90	N	Same as the primary analysis	6.3.2.1
	Supportive	Logistic regression	MITT	Day 5/30 separately	N	Same as the primary analysis	6.3.2.1
	Subgroup (demographic and baseline)	Logistic regression	MITT	Day 90	N	Same as the primary analysis	6.3.2.1
Secondary: Proportion of BI at Day 90 ≥85	Primary	Logistic regression	MITT	Day 90	Y	MI of BI score except for death (worst score assigned)	6.3.2.2
	Sensitivity	Logistic regression	MITT	Day 90	Y	LOCF except for death (worst score assigned)	6.3.2.2

Secondary: mRS Scores at 90 as a Continuous Variable	Primary	MMRM	MITT	Day 5/30/90	Y	death assigned to worst score	6.3.1.3
	Secondary	MMRM	MITT	Day 5/30/90	N	death assigned to worst score	6.3.1.3
Secondary: BI Scores at 90 as a Continuous Variable	Primary	MMRM	MITT	Day 5/30/90	Y	death assigned to worst score	6.3.2.3
	Secondary	MMRM	MITT	Day 5/30/90	N	death assigned to worst score	6.3.2.3
Secondary: Montreal Cognitive Assessment (MoCA) at Day 90 ≥22, 24, 26	Primary	Logistic regression	MITT	Day 90	Y	Single imputation except for death (worst score assigned)	6.3.3.1
	Secondary	Logistic regression	MITT	Day 90	N	Single imputation except for death (worst score assigned)	6.3.3.1
Secondary: MoCA at Days 5/30/90 as repeated measures	Primary	MMRM	MITT	Day 5/30/90	Y	death assigned to worst score	6.3.3.2
	Secondary	MMRM	MITT	Day 5/30/90	N	death assigned to worst score	6.3.3.2
Secondary: Stroke Impact Scale-16 (SIS-16) ≥median at Day 90	Primary	Logistic regression	MITT	Day 90	Y	Single imputation except for death (worst score assigned)	6.3.4.1
	Secondary	Logistic regression	MITT	Day 90	N	Single imputation except for death (worst score assigned)	6.3.4.1
Secondary: SIS-16 at Days 5/30/90 as repeated measures	Primary	MMRM	MITT	Day 5/30/90	Y	death assigned to worst score	6.3.4.2
	Secondary	MMRM	MITT	Day 5/30/90	N	death assigned to worst score	6.3.4.2
Secondary: Excellent Outcomes on NIHSS Scale at Day 90 (2 definitions)	Primary	Logistic regression	MITT	Day 90	Y	MI of NIHSS score except for death due to stroke (worst score assigned)	6.3.5.1
	Secondary	Logistic regression	MITT	Day 90	N	MI of NIHSS score except for death due to stroke (worst score assigned)	6.3.5.1
Secondary: Change from Baseline in NIHSS Scale at Days 5/30/90 as repeated measures	Primary	MMRM	MITT	Day 5/30/90	Y	death due to stroke assigned to worst score	6.3.5.2
	Secondary	MMRM	MITT	Day 5/30/90	N	death due to stroke assigned to worst score	6.3.5.2

Appendix B

Table 1*^: Summary of Bayesian framework of the primary endpoints

	Day 90	Probability of Odds Ratio versus Placebo >1.2	
		Initial Prior Before Study ACTION	Posterior Following Study ACTION
1	Composite of mRs (0,1) and BI ≥95	30%	52%
2	mRs excellent outcome (0, 1)	30%	39%
3	BI excellent outcome ≥ 95	30%	54%

Table 2*^: Summary of Bayesian framework of dichotomous secondary endpoints

	Day 90	Probability of Odds Ratio versus Placebo >1.2	
		Initial Prior Before Study ACTION	Posterior Following Study ACTION
4	mRS score shift	50%	32 %
5	SIS-16 Proportion of subject ⩾ median	50%	47 %
6	MOCA Proportion of subjects with MoCA ⩾26	50%	45 %
7	NIHSS excellent outcome criteria 2 (NIHSS of 0 or 1, or at least an 8-point improvement from Baseline)	50%	30 %

Table 3*^: Summary of Bayesian framework of continuous secondary endpoints

	Day 90	Probability of effect size versus Placebo >0	
		Initial Prior Before Study ACTION	Posterior Following Study ACTION
8	BI score as continuous	50%	65%
9	SIS-16 as continuous	50%	82%
10	MoCA as continuous	50%	81%
11	NIHSS change from baseline as continuous	50%	82%

*: As the functional endpoints mRS and Barthel Index were not primary endpoint of study 101SK201 (ACTION) and the composite of mRS and BI was one of the four related functional endpoints, the findings in ACTION are likely subject to uncertainty beyond the estimated distribution nominally suggested. To account for this additional uncertainly, the prior distribution derived from study 101SK201 (ACTION) will be down-weighted by assuming the sample size of ACTION was 50% of the actual and thereby increasing the variance of the prior distribution. This down-weighting is applied to the Bayesian framework of inference for all primary and secondary efficacy endpoints.

^: For dichotomized endpoints in Table 1 and 2, an "initial prior" probability distribution describing the general likelihood of success in Phase 2 acute ischemic stroke trial based on literature will also be applied. A probability distribution that gives 30% chance for the odds ratio> 1.2 and 14% chance for the odds ratio > 1.45 (assuming a normal distribution of log odds ratio) is utilized based on a study of clinical development success rates 2006-2015 [2]. For table 3, the "initial prior" is a normal distribution with mean 0, and standard deviation 3 times that of the SE of each endpoint from analysis on ACTION data.

References:

[1] Hacke W, et al. "Thrombolysis with Alteplase 3 to 4.5 Hours after Acute Ischemic Stroke". *The New England Journal of Medicine*. 2008. 359(13):1317-1329.

[2] Biotechnology Innovation Organization (BIO), Clinical development success rates 2006-2015. June 2016.

Link:
https://www.bio.org/sites/default/files/Clinical%20Development%20Success%20Rates%202006-2015%20-%20BIO,%20Biomedtracker,%20Amplion%202016.pdf