NOVARTIS

Clinical Development

Afinitor®/Votubia®/Everolimus

CRAD001M2304 / NCT02962414

A three-arm, randomized, double-blind, placebo-controlled study of the efficacy and safety of two trough-ranges of everolimus as adjunctive therapy in patients with tuberous sclerosis complex (TSC) who have refractory partial-onset seizures

RAP Module 3 – Detailed Statistical Methodology – Addendum 1

Author: [redacted]

Document type: RAP Documentation

Document status: Final version 2.0

Release date: March 16, 2016

Number of pages: 98

DOCUMENT CONTENT

This document describes in detail the primary analysis that will be incorporated into the CSR, consisting of double-blind Core phase comparative analyses between everolimus and placebo, plus descriptive analyses of the data from the first part of the Extension phase. The document also describes the analyses for the final analysis, conducted at the end of the trial, focusing on the long-term evaluation of everolimus using all data on everolimus from both the Core and Extension phases of the trial. The study will include interim safety reviews by a Data Management Committee, for which a separate analysis plan will be prepared.

Document History – Changes compared to previous version of RAP Module 3

Version	Date	Changes
1.0	26-Jul-2013	First approval
1.0 – Amendment 1	20-Aug-2015	- Update of the primary endpoint calculations, following protocol amendment 2. Sensory seizures that have been shown to be partial onset with an ictal EEG will be counted in the primary endpoint calculations. - Further clarifications on how to obtain one-sided p-values for the CMH and Rank ANCOVA test statistics are added. - Subgroup analyses of efficacy and safety by number of concomitant AEDs added. - Update definition of the treatment arms to be displayed for the Long-Term Evaluations of safety and efficacy, following the new study design for the extension phase of the trial introduced by the protocol amendment 2. - Addition of two sensitivity analyses for the primary endpoints, to evaluate the potential impact of missing data in the daily seizure diary. - Update of the groupings used to analyze certain types of seizure categories of particular interest. Particularly, sensory seizures are now considered of interest by protocol amendment 2. - Review of the definition of major protocol deviations leading to exclusion from the Per-protocol Set: - Criteria for minimum duration of baseline changed from 54 days to 49 days - Clarifications on the major PDs related to concomitant AED regimens and everolimus treatment received. - Removal of the criteria for the use of rescue medication - Vineland data: - Further clarifications added on the process to obtain the subdomain/domain and global scores using an external software. - Due to improper filling-out of the questionnaires, subdomain scores were missing frequently. An attempt is being made to retrospectively collect the missing information. Inclusion of a supportive analysis using all available data is described in this RAP amendment. - Raw scores from the Wechsler Non Verbal scale will be used, converted to a z-score. No attempt is made to derive standardized T-scores comparing to the reference population provided with the WNV evaluation tool.

Version	Date	Changes
		- Columbia Suicide Severity Rating Scale: clarification on the analysis of incomplete calls to the IRT when followed by a complete call. - Update of the PK/PD analyses - Update of the growth data analyses to use the references provided by the World Health Organization (WHO), rather than the references provided by the Center for Disease Control (CDC), (when WHO references are available). - Further clarifications on reporting of prior Antiepileptic drugs received by patients. - Further clarifications added on the use of the CTCAE v4.03 for the derivation of hematological/biochemical parameter grades. - SAS procedure to estimate the mean difference (and its 95% CI) between two continuous variables is added. - Further minor updates, clarification and corrections
2.0 - Amendment 2	18-Nov-2015	The changes implemented in this CSR RAP-Amendment 2 are the following: Changes related to efficacy analysis: Addition of two subgroup analyses: 1) Response rate and percentage reduction in seizure frequency in the subgroup of patients with at least one Subpendymal Astrocytoma lesion (SEGA). 2) Response rate and percentage reduction in seizure frequency by considering only the Generalized Tonic Clonic Seizures (GTCS) (seizure code: IC4). For the QOLIE-AD-48 and the QOLIE-31-P questionnaires, the rule for the derivation of the total scores in case of missing subscales will be updated in the RAP Module 3 in order to be consistent with the Scoring Manual. Changes related to safety analysis: The term “Clinically Notable Adverse Event (CNAE)” will be replaced by “Adverse Event of Special Interest (AESI)”. Other general changes: Minor updates, clarification and corrections added throughout the documents. [redacted]

Version	Date	Changes
		[redacted] All other updates, clarifications and corrections arose during the review of dry runs and were requested in order to meet the requirements of submission documents (Case Study Report for safety updates) or to correct inconsistencies between RAP module 3 and RAP module 7.
2.0 - Addendum 1	16-Mar-2016	The changes implemented in this CSR RAP-Addendum 1 are the following: - Details on derivation of projected Cmin for PK analyses were added (section 3.13.1.2). - Two figures to assess the effect of everolimus in seizure frequency were added: waterfall plot and cumulative plot of percentage reductions from baseline in seizure frequency. - [redacted] - The term “Clinically Notable Adverse Event” (CNAE) was replaced by “Adverse Event of Special Interest” (AESI). - Dose-proportional tests by age range and by the use of CYP3A4 inducer/inhibitor (without inducer and inhibitor, with inducer and without inhibitor, with inhibitor and without inducer) will be addressed in the CSPD documents and removed from this statistical analysis plan. - For the FDA primary endpoint using a rank ANCOVA test, the effect size (measuring the magnitude of the everolimus effect) will be assessed by using a stratified bootstrap method to estimate the difference between median percentage reduction from baseline between everolimus treatment arms and placebo.

Statistical methods planned in the protocol and determination of sample size

Data will be analyzed by Novartis and/or a designated CRO according to the data analysis Section 10 of the study protocol which is available in Appendix 16.1.1 of the CSR. Important information is given in the following sections and details are provided, as applicable, in Appendix 16.1.9 of the CSR.

Statistical and analytical plans

The planned analysis is described in Section 10 of the protocol (Appendix 16.1.1).

Categorical data will be presented as frequencies and percentages. For continuous data, mean, standard deviation, median, minimum, and maximum will be presented unless specified in a different way.

It is planned that the data from all centers that participate in this protocol will be used. Primary efficacy analysis will be stratified, reflecting the stratified nature of the randomization method.

Subjects and treatments

Analysis sets

Full Analysis Set (FAS): The FAS comprises all patients to whom study treatment has been assigned by randomization. Following the intent-to-treat principle, patients will be analyzed according to the treatment arm and stratification factor that they were assigned to at randomization.

Safety Set: The Safety Set includes all patients who received at least one dose of study medication in the Core phase, and had at least one post-baseline safety assessment in the Core phase (where the statement that a patient had no adverse event (on the Adverse Events eCRF) constitutes a safety assessment).

There are three treatment groups in the study: everolimus titrated to a trough concentration of 3 to 7 ng/ml, everolimus titrated to a trough concentration of 9 to 15 ng/ml, and placebo. The usual approach for the Safety Set would be to analyze patients according to study treatment received. However, that is not possible because the two everolimus arms have the same treatment received (i.e., everolimus), and are only differentiated by differing targeted trough ranges. Therefore, patients randomized to one of the two everolimus arms and who received at least one dose of everolimus, will be analyzed in the Safety Set according to the everolimus treatment arm assigned at randomization. Similarly, patients randomized to placebo and who received at least one dose of placebo, will be analyzed in the placebo arm of the Safety Set.

That is, only patients who received the wrong treatment regimen throughout their entire time in the Core phase will be reassigned to a different treatment arm in the Safety Set compared with the FAS, as follows:

- Patients randomized to either of the two everolimus arms and who received only placebo tablets during the Core phase will be reassigned to the placebo arm in the Safety Set. This could happen for example if the first kit of study medication contained only placebo tablets (i.e., given in error), and the patient took at least one tablet and then subsequently discontinued before taking any everolimus tablets.
- Patients randomized to the placebo arm and who received only everolimus tablets during the Core phase will be reassigned to the everolimus 3 to 7 ng/ml arm in the Safety Set. Again this would require a dispensing error in the first kit of study medication, followed by early discontinuation. Note that the reassignment is to the lower everolimus targeted trough arm, chosen because such an error in starting dose would be recognized via IRT before any dose increase could be made.

Note that there will be no attempt to reassign patients from one everolimus arm to the other everolimus arm based on observed everolimus concentrations during the Core phase. For example, a patient randomized to the everolimus 9 to 15 ng/ml arm and whose observed everolimus trough concentrations ranged from 3 to 7 ng/ml during the entire Core phase would still be considered to have treatment received = everolimus 9 to 15 ng/ml. This is because increases in the dose of everolimus would have been recommended during the Core phase in order to reach the targeted trough range of 9 to 15 ng/ml, and any associated toxicity should therefore be attributed to the 9 to 15 ng/ml arm and not the 3 to 7 ng/ml arm.

Per-Protocol Set (PPS): The PPS will be a subset of the patients in the FAS who are compliant with requirements of the protocol, who are evaluable for efficacy and who have completed a minimum exposure requirement. However, if a patient discontinued for lack of efficacy, adverse event or death before the minimum exposure requirement could be met, or before he/she could become evaluable for efficacy, that patient will still be included in the PPS.

Patients will be evaluable for efficacy if their seizure diary continues until at least day 70 of the Core phase, i.e., at least the first 4 weeks of the maintenance period of the Core phase.

The minimum exposure requirement is defined as receiving study treatment on ≥ 50% of days during the first 10 weeks of the Core phase.The following list of major protocol deviations will constitute reason for excluding the patient from the PPS:

- Absence of diagnosis of TSC or partial-onset epilepsy.
- Previously failed < 2 sequential regimens of single or combined AEDs.
- Prior therapy with systemic or topical mTOR inhibitors within the exclusion criteria time windows.
- Baseline seizure diary with < 16 ESC-approved partial-onset seizures.
- Baseline seizure diary containing a seizure-free period of ≥ 21 consecutive days.
- Baseline seizure diary < 49 days in duration.
- Did not receive 1-3 AEDs at same dose from 4 weeks prior to screening visit to baseline visit
- Change in dose for more than 7 days or in number of concomitant AEDs during Core Phase or interruption of AED > 7 days
- Missing seizure information on > 50% of days in the Core phase or the Baseline phase.
- Received the wrong treatment regimen throughout their entire time in the Core phase, defined as:
    - Patients randomized to either of the two everolimus arms and who received only placebo tablets during the Core phase.
    - Patients randomized to the placebo arm and who received only everolimus tablets during the Core phase.

Long-Term Evaluation Efficacy Set (LTE Efficacy Set): The LTE Efficacy Set is intended to capture all data on everolimus in the trial, from both the Core phase and the Extension phase. It consists of all patients who received at least one dose of everolimus in the trial and

have at least one post-baseline efficacy measure. Patients will be analyzed as a single everolimus group as well as separately according to their randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3-7 ng/ml and 9-15 ng/ml, or placebo). This analysis set will be used for long-term efficacy analyses of everolimus.

Before protocol amendment 2, patients from the two everolimus groups were to continue in the Extension phase with the same targeted trough range from Core phase, while placebo patients were to be switched to receive everolimus during the Extension phase with a 3-15 ng/ml targeted trough. Protocol amendment 2 introduced the possibility for investigators to make the dose adjustments in the Extension phase based on the Cmin value. If investigators choose not to modify dosing themselves, the dose was determined by the IRT system to maintain the everolimus trough concentration range between 3-15 ng/ml for all patients. In order to avoid compromising the blinded nature of the study in the earlier Core phase (including the investigator, patient and sponsor), all patients were to be transitioned towards a common targeted trough range of 6 to 10 ng/ml before any Cmin value is shared with the investigator. For this reason, for analyses of long-term evaluation data from the Core and Extension phase, patients will be analyzed as a single everolimus group ("All patients") as well as separately according to their randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3-7 ng/ml and 9-15 ng/ml and placebo). Note that "placebo" arm will only contain data from the Extension phase when patients received everolimus.

Long-Term Evaluation Safety Set (LTE Safety Set): The LTE Safety Set is intended to capture all data on everolimus in the trial, from both the Core phase and the Extension phase. It consists of all patients who received at least one dose of everolimus in the trial and have at least one post-baseline safety measure. Patients will be analyzed as a single everolimus group as well as separately according to their randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3 to 7 ng/ml and 9 to 15 ng/ml, or placebo. This analysis set will be used for long-term safety analyses of everolimus.

For the same reason as described above for the LTE Efficacy Set, patients will be analyzed as a single everolimus group ("All patients") as well as separately according to the randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3 to 7 ng/ml and 9 to 15 ng/ml and placebo). Note that "placebo" arm will only contain data from the Extension phase when patients received everolimus.

Pharmacokinetic analysis set: There will be no formal pharmacokinetic analysis set. The pharmacokinetic analyses will be performed in the Safety Set and the LTE Safety Set using all available PK samples from the Confirmed PK Sample Set or PK Sensitivity Sample Set, as defined below (section 3.13.1.1).

Patient demography and other baseline characteristics
Demographic (age, gender, ethnicity, weight, height, BMI, BSA) and disease characteristics (time since diagnosis of partial onset seizures, prior epilepsy surgery, prior VNS, prior ketogenic diet, seizure history (status epilepticus, simple partial seizure, complex partial seizure, secondarily generalized tonic clonic convulsion, primarily generalized tonic clonic convulsion, absence, myoclonus, tonic, clonic, other), major/minor features of TSC) will be listed and summarized descriptively by treatment group in the FAS and in the LTE Efficacy

Set. Prior anti-TSC therapies and number and type of prior antiepileptic drugs (AEDs) will also be listed and summarized by treatment group.

Treatment (study drug, background AED regimen, other concomitant therapies)
Duration of study treatment exposure, cumulative dose and dose intensity will be summarized by treatment group. The numbers of patients with dose changes/interruptions will be presented by treatment group, along with reasons for the dose change. These analyses will be conducted over the Core phase on the Safety set, and over the entire study using the LTE Safety Set.

Concomitant medications and significant non-drug therapies taken concurrently with the study drugs will be listed and summarized by ATC term, preferred term, and treatment arm by means of frequency counts and percentages. Of particular interest will be the medications used in the background AED regimen. In addition, specific tables will be produced for rescue medication and medications that may affect everolimus mechanism of action (i.e., CYP3A4, PgP).

Efficacy evaluation

Analysis of primary efficacy endpoint

The primary efficacy endpoint is the reduction from baseline in the frequency of partial-onset seizures during the maintenance period of the Core phase. Two separate definitions of the primary efficacy endpoint will be used, owing to the differing regulatory preferences between Europe and USA for demonstrating efficacy of an antiepileptic medication. Whereas the EMA prefers **response rate** to be the primary variable, the FDA prefers **percentage reduction in seizure frequency**. It is understood that each Agency will use their preferred variable as the primary variable, with the other (non-primary) variable being used in a supportive analysis. As each Agency will only use their preferred primary variable to make a decision on the primary objective, the full alpha level can be used for each Agency's primary variable, without correction for multiplicity. This approach has been endorsed by both EMA and FDA.

Response rate and percentage reduction in seizure frequency will be determined using counts of partial-onset seizures, based on seizure diaries that are completed by the patient or caregiver throughout the trial. During the baseline phase the investigator will review the known seizure types of each patient with the Epilepsy Study Consortium (ESC), an independent group of experienced epileptologists charged with harmonizing the classification of partial-onset seizures in the trial. Only seizures approved by the ESC and agreed to by the investigator will be entered into the eCRF and counted as partial-onset seizures.

The following definitions are required to calculate response rate and percentage reduction in seizure frequency:

1. Average weekly seizure frequency in the Baseline phase (SF_B) = 7 × number of partial-onset seizures recorded over the 8 week prospective Baseline phase ÷ number of non-missing seizure diary days in the 8 week prospective Baseline phase.
2. Average weekly seizure frequency in the maintenance period of the Core phase (SF_M):

- If patient does not discontinue during the 6 week titration phase, $SF_M = 7 \times$ number of partial-onset seizures recorded during the maintenance period of the Core phase ÷ number of non-missing seizure diary days in the maintenance period of the Core phase.
- Otherwise, $SF_M = 7 \times$ number of partial-onset seizures recorded during the titration period of the Core phase ÷ number of non-missing seizure diary days in the titration period of the Core phase.

That is, patients who discontinue prior to the maintenance period have seizure frequency determined using their data from the titration period, thereby assuring that all patients with seizure data in the Core phase have a value for SF_M.

3. Percentage reduction from baseline in average weekly seizure frequency during the maintenance period of the Core phase (%Red) = $100 \times (SF_B - SF_M) \div SF_B$.

Point 3 is the FDA primary variable, percentage reduction in seizure frequency, defined by %Red.

A responder is a patient with ≥ 50% reduction from baseline in average weekly partial-onset seizure frequency during the maintenance period of the Core phase, that is, when %Red ≥ 50. Response rate is the percentage of responders in a treatment group.

Response rate will be compared between each everolimus arm versus the placebo arm in the FAS, using Cochran-Mantel-Haenszel (CMH) chi-square tests stratified by age subgroup. The Bonferroni-Holm procedure will be used to ensure an overall family-wise Type I error rate of 2.5% one-sided. Response rates will be provided with exact 95% confidence intervals, and the odds ratio for each everolimus arm versus placebo will be obtained from logistic regression models stratified by age subgroup.

Percentage reduction in seizure frequency, the FDA primary variable, will be compared between each everolimus arm versus the placebo arm in the FAS using rank ANCOVA, with baseline average weekly seizure frequency as a covariate, and stratified by age subgroup. The Bonferroni-Holm procedure will be used as a multiplicity correction to ensure an overall family-wise Type I error rate of 2.5% one-sided. The median percentage reduction from baseline will be presented for each treatment group, along with 95% bootstrap confidence intervals.

For the EMA primary variable of response rate, a supportive analysis will be the FDA primary analysis on percentage reduction in seizure frequency, and vice versa.

The primary analysis of both the EMA and FDA primary variables will be repeated in the Per Protocol Set.

Descriptive statistics (n, mean, median, standard deviation, minimum and maximum) for average weekly seizure frequency in the Baseline phase (SF_B), in the maintenance period of the Core phase (SF_M), and percentage reduction from baseline (%Red), will be presented by treatment group in the FAS. The quantities for seizure frequency will also be summarized in units of 28 days instead of weekly, i.e., average number of partial-onset seizures per 28 days in the Baseline phase, and per 28 days in the maintenance period of the Core phase.

In addition, sensitivity analyses will be conducted in the FAS to assess robustness of the primary analysis to (i) patient discontinuation before the end of the Core phase, (ii) the

calculation of seizure frequency across the entire Core phase (i.e., titration period as well as maintenance period), (iii) missing daily seizure data.

For (i), the sensitivity analysis will assume no change from baseline in seizure frequency for patients discontinuing in the Core phase, compared with the approach for the primary analysis where the actual seizure frequency calculated up to the day of discontinuation will be used. Response rate and percentage change from baseline in seizure frequency will be recalculated for this sensitivity analysis and analyzed per the primary analysis.

For (ii), the sensitivity analysis will use the average weekly seizure frequency over the entire Core phase, that is, over the 6-week titration period plus the 12-week maintenance period. This compares with the primary analysis where titration period data will not be used to calculate average weekly seizure frequency if the patient continues into the maintenance period. The variable to be used in the sensitivity analysis, labeled SF_{TM} is defined as follows:

- Average weekly seizure frequency in the entire Core phase (SF_{TM}) = 7 × number of seizures recorded over the entire Core phase (titration plus maintenance) ÷ number of non-missing seizure diary days in the entire Core phase (titration plus maintenance).

Response rate and percentage reduction in seizure frequency will be re-calculated using SF_{TM} instead of SF_M, and analyzed as per the primary analysis.

For (iii), two approaches will be considered:

- Best case scenario: this sensitivity analysis will assume that patient experienced no seizures (i.e. missing data imputed to 0) in the maintenance period on the days where the data is reported as missing in the eCRF. For the baseline phase, it will be assumed that the patient experienced the maximum number of seizures that she/he has experienced during the baseline phase (i.e. missing data imputed to maximum daily seizure count).

- Worst case scenario: this sensitivity analysis will assume that the patient experienced the same maximum number of seizures that she/he experienced during the maintenance period on the days where the data is reported as missing in the eCRF in the maintenance period (i.e. missing data imputed to maximum daily seizure count). For the baseline phase, it will be assumed that the patient experienced no seizures (i.e. missing data imputed to 0) on these missing days.

Response rate and percentage change from baseline in seizure frequency will be recalculated for this sensitivity analysis and analyzed per the primary analysis.

Subgroup analyses will be performed to compare each everolimus arm versus placebo on the primary analyses of response rate or percentage reduction in seizure frequency. The subgroups that will be analyzed include gender, age, race and ethnicity. Subgroup analyses of the main efficacy and safety endpoints will be performed on patients randomized in Japan, for use by the Japanese regulatory authority.

Analysis of secondary endpoints

Seizure freedom is defined as remaining seizure-free during the maintenance period of the Core phase (or during the titration period for patients who discontinue during the titration period). The seizure-free rates for each treatment arm in the FAS will be presented along with exact 95% confidence intervals, and the odds ratio for each everolimus arm versus placebo

will be derived from logistic regression models stratified by age subgroup. A sensitivity analysis will be performed where patients who discontinue during the Core phase are assumed not to be seizure-free even if no seizures had been reported before they discontinued.

The **proportions of patients with at least 25% reduction in seizure frequency** during the maintenance period of the Core phase (or titration period for patients who discontinue in the titration period) will be presented in each treatment arm along with exact 95% confidence intervals, and odds ratios for each everolimus arm versus placebo from logistic regression models stratified by age subgroup.

The **distribution of reduction from baseline in seizure frequency** will be categorized into the following six levels using the variable %Red defined above: ($\leq$ -25% (exacerbation); > -25% to < 25% (no change); $\geq$ 25% to < 50% (25% response); $\geq$ 50% to < 75% (50% response); $\geq$ 75% to < 100% (75% response); 100% (seizure freedom)), and the proportions of patients in each category will be presented for each treatment arm.

The frequency of **seizure-free days** per 28 days during the maintenance period of the Core phase will be obtained for each patient, calculated as 28 × number of partial-onset seizure-free days in the maintenance period of the Core phase ÷ number of non-missing seizure diary days in the maintenance period of the Core phase. For patients who discontinue in the titration period, the number of seizure-free days per 28 days during the titration period will be used. A similar quantity will be calculated for the Baseline phase. Then the change from baseline in frequency of seizure-free days per 28 days will be summarized by treatment arm (mean, standard deviation, range). Mean differences between each everolimus arm and the placebo arm in change from baseline in frequency of seizure-free days will be presented, along with 95% confidence intervals.

Treatment duration is defined as the time from randomization until the date of permanent study treatment discontinuation (for any reason) at any time during the Core phase (i.e., titration period or maintenance period). Patients who complete the Core phase without discontinuing will have treatment duration censored on the last day of the Core phase. The treatment duration distributions in each arm will be presented descriptively in the FAS using Kaplan-Meier curves, from which summary statistics will be determined, including the median treatment duration and the proportions of patients still on treatment at 6 and 12 weeks. These statistics will be given as point estimates with 95% confidence intervals. The hazard ratio (and two sided 95% confidence interval) for each everolimus arm versus placebo will be obtained from a Cox proportional hazards model stratified by age subgroup.

A **long-term evaluation of efficacy over the extension phase** will use descriptive statistics of percentage reduction from baseline in partial-onset seizure frequency, response rate and seizure-free days, computed by time interval using the LTE Efficacy Set. For patients randomized to one of the two everolimus arms, all data from the Core phase and the Extension phase will be used, whereas for patients randomized to placebo, only the Extension phase data will be included. There will be 4 different treatment groups displayed: everolimus 3 to 7 ng/ml, everolimus 9 to 15 ng/ml, placebo (i.e., extension phase data from patients previously randomized to placebo) and an overall everolimus arm.

Pharmacokinetic evaluations

PK analyses will be based on evaluable samples from patients in the Safety Set, and for the long-term evaluation of everolimus, in the LTE Safety Set. The validity of PK samples will be confirmed by checking the sampling time window and whether the patient vomited within 4 hours of taking study drug. Only confirmed PK concentrations will be used in the analyses.

For each everolimus treatment group, everolimus concentrations at trough (C_{min}) will be summarized by descriptive statistics at each PK sampling time during the Core phase in the Safety Set, and during the Core and Extension phases in the LTE Safety Set. The descriptive statistics will be arithmetic and geometric mean, median, standard deviation, coefficient of variation (CV), geometric CV, minimum and maximum.

To evaluate the impact of CYP3A4/PgP inducer and inhibitor on everolimus PK, a linear mixed model will be fitted to log-transformed C_{min} concentration at steady state including log transformed dose, use of CYP3A4/PgP inducer and inhibitor and other factors as appropriate as covariates.

To explore the relationship between everolimus C_{min} and the response rate as defined for the EMA primary analysis, a logistic regression model will be used to fit the response data using time-normalized C_{min} levels in the maintenance period of the Core phase, stratifying by age subgroup and adjusting for additional risk factors if appropriate.

A linear regression model will be used to characterize the impact of exposure on the average weekly seizure frequency. The model will include baseline seizure frequency (SF_B) and time-normalized C_{min} in the maintenance period of the Core phase, both in log scale.

A linear mixed model with repeated measurements will be used to link the post baseline average weekly seizure frequency to the time normalized C_{min} in defined time intervals during the Core phase. The model will be adjusted by the baseline frequency of seizure (SF_B). Other covariates will be included if appropriate.

The relationship between everolimus C_{min} and selected safety endpoints will also be explored by an appropriate model-based approach. In addition the time to first event of selected AEs will be fitted with an extended Cox model with projected C_{min} as a time varying covariate, stratified by age subgroups.

To evaluate the effects of everolimus on the exposure of 12 different AEDs, descriptive statistics will be used to compare the pre-dose concentrations of AEDs at Visits 1 and 2 (AED alone) with those at Visits 3 and 4 (AED plus everolimus) in the two everolimus treatment groups. In addition, separate linear mixed models will be fitted to log transformed concentration levels of each of the 12 AEDs, including period (before and after everolimus administration) as a fixed effect and patient as random effect. Geometric mean ratios of the concentrations with and without everolimus (as reference) and their 90% confidence intervals will be calculated from the model.

Safety evaluation

The assessment of safety will be based mainly on the frequency of adverse events and on the number of laboratory values that fall outside of pre-determined ranges. Other safety data

including vital signs, neuropsychological scales and the Columbia Suicide Severity Rating Scale will also be considered.

Safety analyses will be performed on the Safety Set, and on the LTE Safety Set.

Safety summary tables will include only assessments collected no later than 30 days after study treatment discontinuation. All safety assessments will be listed and those collected later than 30 days after study treatment discontinuation will be flagged.

Serious AEs, AEs leading to discontinuation, AEs requiring dose reduction/interruption, related AEs and notable AEs will be summarized by system organ class, preferred term and treatment. Maximum grade by AEs will be also summarized. The time to first occurrence of selected AEs will be assessed in the double blind Core phase.

All laboratory values will be converted into SI units. Shift tables using CTCAE grades when available or low/normal/high/ (low and high) classification when no CTCAE grades are available will be used to compare baseline to the worst on-treatment value.

The **Vineland-II** instrument (Vineland Adaptive Behavior Scales 2nd Edition) will be completed by patients at baseline, end of Core phase, and then every 6 months in the Extension phase plus at the end of treatment in the Extension phase. Changes from baseline in domain and subdomain scores will be listed and summarized descriptively by treatment group in both the Safety Set and the LTE Safety Set.

The **Wechsler Nonverbal** instrument (WNV: Wechsler Nonverbal Scale of Ability) is for patients aged 4-21 years at baseline, and will be completed at baseline, end of Core phase, and then every 6 months in the Extension phase plus at the end of treatment in the Extension phase. Changes from baseline in overall and test scores will be listed and summarized descriptively by treatment group.

The **Columbia Suicide Severity Rating Scale** will be completed by patients at all visits (except weeks 1, 3, 5 and 19 when only PK samples are collected, and the End of Study visit). The proportions of patients in each treatment group in the Safety Set with suicidal ideation, suicidal behavior and suicidal ideation or behavior, will be presented by treatment group. This analysis will be repeated using the LTE Safety Set.

Interim analyses

No interim analysis for efficacy is planned, but there will be DMC reviews of the ongoing safety data as the trial progresses. The first meeting will review data from the first 6 months of the study, with subsequent meetings every 6 months thereafter. As outlined in the DMC charter, the safety review outputs will be prepared by an independent statistician and an independent programmer from Novartis, neither of whom will belong to the trial team or will be involved in any other aspects of the trial conduct. Semi-blinded results will be shared with the DMC members using a secured web portal to which Novartis personnel have no access (except independent statistician/programmer). After the meeting, the DMC chair will provide the recommendation to Novartis Oncology global development head. Apart from the recommendation made, the Novartis team members will remain blinded to any study results.

Other topics

Patient reported outcome

Quality of life will be assessed using three age-specific questionnaires: the Quality of Life in Childhood Epilepsy (QOLCE) for patients aged ≤ 10 years, the Quality of Life in Epilepsy Inventory for Adolescents-48 (QOLIE-AD-48) for patients aged 11 to ≤ 17 years, and the Quality of Life in Epilepsy Inventory-31-Problems (QOLIE-31-P) for patients aged ≥ 18 years. Descriptive statistics for the overall quality (QOL) score and subscale scores will be presented by time point and treatment group in the FAS for each questionnaire.

Growth data

Growth data will only be analyzed as part of the long-term evaluation of everolimus among patients under the age of 18, i.e., using patients from the LTE Safety Set who were under 18 years of age on the start date of everolimus.

Data will be summarized descriptively for each treatment group at each time point. The data will consist of height, height velocity, weight and weight velocity. In addition, based on the data collected during the study and published reference information, the standard deviation score (SDS, also called z-score) will be computed for each patient at each time point for height, height velocity, weight and weight velocity.

Descriptive statistics of these endpoints will be presented by time point and the z-scores will allow identification of potential outliers. Growth velocity during the trial will also be compared with growth velocity at baseline (if sufficient pre-baseline data are available).

Comparison of height at 18 years with mid-parental heights will be performed, by presenting the number of patients with a height higher/lower than expected.

Puberty stage

Puberty stage will only be analyzed as part of the long-term evaluation of everolimus on the LTE Safety Set.

Tanner Stage includes two components for boys, namely testis and pubic hair, and two components for girls: breast development and pubic hair. It is expected that data will become available during the trial on a proportion of patients as they go through puberty attaining higher levels of the Tanner Stage. For the age at which Tanner Stages 2-5 are achieved, age at thelarche (females), age at menarche (females) and age at adrenarche (males), summary statistics from Kaplan-Meier distributions will be determined, including the median age and the proportions of patients reaching these milestones at some given ages. These statistics will be given as point estimates with 95% confidence intervals. The percentage of patients who will reach Tanner Stage 2 during the study will be calculated among the number of patients with Tanner Stage 1 at Baseline (prior to the start of everolimus). The percentage of patients who will reach Tanner Stage 3 during the study will be calculated among the number of patients with Tanner Stage 1 or 2 at Baseline.

Similar rules will be applied for the age at Tanner Stages 4 and 5. Age at thelarche, age at menarche and age at adrenarche will be assessed among patients who have not yet reached these development milestones at baseline.

Potential delayed puberty in girls is defined as failure to attain Tanner Stage 2 (for both breast development and pubic hair) by age 13, or absence of menarche by age 15 or within 5 years of attainment of Tanner Stage 2 (Fenichel et al. 2012). Potential delayed puberty in boys is defined as failure to attain Tanner Stage 2 (for both testis and pubic hair) by age 14 (Crowley et al. 2012). Rates of potential delayed puberty will be presented for boys and girls separately, along with 95% confidence intervals, on the population at risk of delayed puberty at baseline. Potential cases identified through this algorithm, will be then clinically reviewed by assessing all available information in order to conclude the clinical relevance of the delay.

TSC genetic testing

Proportions of patients with TSC1 and TSC2 gene mutations will be presented per treatment arm in the FAS.

Determination of sample size

The sample size was chosen to provide adequate power for the primary objective comparing seizure frequency between each everolimus arm and the placebo arm. The sample size calculation provided here is based exclusively on response rate, the EMA primary endpoint, but it is also expected to provide sufficient patients for the power of the FDA primary endpoint, percentage reduction in seizure frequency. This is because response rate is a binary transformation of percentage reduction in seizure frequency, and therefore likely to be less sensitive owing to the loss of information going from a continuous variable to binary.

It was assumed that response rates would be 15% in the placebo arm and 35% in each of the two everolimus arms. That is, there was no *a priori* strong expectation that the higher targeted trough everolimus arm 9 to 15 ng/mL would deliver a higher response rate than the lower targeted trough everolimus arm 3 to 7 ng/mL, as better efficacy may be mitigated by worse tolerability. For this reason, the testing strategy was to simultaneously compare each pairwise comparison, splitting the significance level, rather than testing hierarchically starting with the higher trough arm for example.

Using nQuery version 6.1 it was determined that a sample size of 355 patients would ensure 90% power for each of the primary comparisons of each everolimus arm versus placebo, assuming one-sided 1.25% significance levels for each CMH chi-square test, and assuming balanced randomization (i.e., 115 patients per randomization arm). Due to a mistake in the

IRT system discovered early in the trial, preventing dose titrations despite C_{min} values outside the targeted trough range, it was decided to increase the sample size in the everolimus 9 to 15 ng/ml arm by 10 patients, i.e. 125 patients in total.

Table of contents

List of figures

List of tables

Clinical Study Report - Appendix 16.1.9 Documentation of statistical methods

List of abbreviations

AE	Adverse Event
AED	Antiepileptic Drug
ANCOVA	Analysis of Covariance
ATC	Anatomic Therapeutic Chemical (Classification)
BMI	Body Mass Index
BSA	Body Surface Area
CMH	Cochran-Mantel-Haenszel
C_{min}	Pre-dose concentration of drug at steady-state
$C_{min,TN}$	Time normalized C_{min}
CRO	Contract Research Organization
C-CASA	Columbia Classification Algorithm for Suicide Assessment
C-SSRS	Columbia Suicide Severity Rating Scale
CTC	Common Terminology Criteria
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
DAR	Dosage Administration Record
DI	Dose Intensity
DMC	Data Monitoring Committee
DRF	Diagnostic Review Form
eCRF	Electronic Case Report/Record Form
eC-SSRS	Electronic Columbia Suicide Severity Rating Scale
EEG	Electroencephalography
EMA	European Medicines Agency
ESC	Epilepsy Study Consortium
FAS	Full Analysis Set
FDA	Food and Drug Agency
HLT	High Level Term
IRT	Interactive Response Technology
IVRS	Interactive Voice Response System
LTE	Long-Term Evaluation
MAP	Master Analysis Plan
MedDRA	Medical Dictionary for Regulatory Activities
NA	Not Applicable
NOS	Not Otherwise Specified
PgP	P-glycoprotein
PK	Pharmacokinetic
PPS	Per Protocol Set
PT	Preferred Term
QOL	Quality of Life
QOLCE	Quality of Life in Childhood Epilepsy
QOLIE-31-P	Quality of Life in Epilepsy Inventory – 31 – Problems
QOLIE-AD-48	Quality of Life in Epilepsy Inventory for Adolescents – 48

RAP	Report Analysis Plan
SAE	Serious Adverse Event
SD	Standard Deviation
SDS	Standard Deviation Score
SEC	Safety Event Category
SI	Standard International
SIF	Seizure Identification Form
SOC	System Organ Class
TEAE	Treatment-Emergent Adverse Event
TSC	Tuberous Sclerosis Complex
VAP	Validation and Planning
VNS	Vagal Nerve Stimulator
WHO	World Health Organization
WNV	Wechsler Nonverbal

1 Introduction

This Reporting and Analysis Plan (RAP) Module 3 incorporates the latest project standards from the Master Analysis Plan (MAP), i.e., "Module 3 – Detailed Statistical Methodology – Amendment 4", finalized on 06-Jan-2010.

Overview of study design and objectives

This is a prospective, multi-center, randomized, double-blind, parallel-group, placebo-controlled, 3-arm Phase III study comparing the efficacy and safety of two trough-ranges of everolimus to placebo as adjunctive therapy in patients with tuberous sclerosis complex (TSC) who have refractory partial onset seizures. A total of 355 patients will be randomized in a 1: 1:1.09 ratio to receive 1) placebo, or 2) everolimus titrated to a trough concentration of 3 to 7 ng/ml or 3) everolimus titrated to a trough concentration of 9 to 15 ng/ml, with all patients remaining on a stable background antiepileptic drug (AED) regimen (of 1-3 AEDs). Randomization will be stratified into 4 groups according to age at baseline: < 6 years, 6 to < 12 years, 12 to < 18 years, and ≥ 18 years. The primary objective is to compare the reduction from baseline in frequency of partial-onset seizures between each everolimus arm versus the placebo arm.

The study includes a prospective Baseline phase during which patients complete a seizure diary for a total of 8 weeks. At the end of the Baseline phase, eligible patients will be randomized and enter an 18-week double-blind Core phase. The Core phase starts with a 6 week titration period, during which up to 3 dose adjustments can be made in order to reach the targeted everolimus trough range, followed by a 12-week maintenance period. Patients continue to record occurrence of seizures on each day throughout the Core phase. After completing the maintenance period of the Core phase, patients will be offered to continue in an Extension phase, where all patients will receive everolimus; this Extension phase allows for a long-term evaluation of the safety and efficacy of everolimus.

The primary efficacy endpoint is the reduction from baseline in the frequency of partial-onset seizures during the maintenance period of the Core phase. Two separate definitions of the primary efficacy endpoint will be used, owing to differing regulatory preferences between Europe and USA for demonstrating the efficacy of an antiepileptic medication. Whereas the European Medicines Agency (EMA) prefers response rate to be the primary variable, the Food and Drug Agency (FDA) prefers percentage reduction in seizure frequency. However, since a response is defined as ≥ 50% reduction from baseline in seizure frequency, the two variables are very similar. It is understood that each Agency will only use their preferred primary variable to make a decision on the primary objective, and therefore the full alpha level will be used for each variable, without correction for multiplicity.

The EMA primary analysis will compare response rates between each everolimus arm versus the placebo arm in the Full Analysis Set (FAS), using Cochran-Mantel-Haenszel (CMH) chi-square tests stratified by age subgroup. In order to accommodate the simultaneous testing of each everolimus arm versus placebo, the Bonferroni-Holm procedure will be used to ensure an overall family-wise Type I error rate of 2.5% one-sided.

The FDA primary analysis will compare percentage reduction in seizure frequency between each everolimus arm versus the placebo arm in the FAS, using rank analysis of covariance (ANCOVA), with baseline average weekly seizure frequency as a covariate, and stratified by age subgroup. Again, the Bonferroni-Holm procedure will be used as a multiplicity correction to ensure an overall family-wise Type I error rate of 2.5% one-sided.

The trial was designed to provide adequate power for the primary objective comparing seizure frequency between each everolimus arm and the placebo arm. With respect to the EMA primary endpoint of response rate, the sample size of 115 patients per arm provides 90% power assuming response rates of 15% on placebo and at least 35% on either everolimus arm. This sample size is also expected to provide at least 90% power for the FDA primary endpoint, percentage reduction in seizure frequency, because response rate is expected to be less sensitive since there is a loss of information going from a continuous variable to binary.

The secondary objectives are to compare each of the two everolimus trough ranges versus placebo with respect to:

- Ability to completely suppress partial-onset seizures
- Proportion of patients with ≥ 25% reduction from baseline in seizure frequency
- Distribution of reduction from baseline in seizure frequency
- Seizure-free days
- Treatment duration
- Quality of life

Further secondary objectives are:

- To assess everolimus in relation to neurocognitive, neurobehavioral and neuro-developmental measures using Vineland-II and Wechsler Nonverbal scales
- To assess the relationship between everolimus concentration at trough and efficacy / safety endpoints
- To evaluate the impact of everolimus on the pre-dose exposure of background AEDs
- To evaluate the effect of the two everolimus trough ranges on long-term seizure reduction
- To evaluate the safety and tolerability of each everolimus trough range in the study population
- To evaluate the impact of everolimus on the risk of suicide using the Columbia Suicide Severity Rating Scale (C-SSRS)

[redacted]

A Data Monitoring Committee (DMC) will review the ongoing safety data as the trial progresses. The first meeting will review data from the first 6 months of the study, with subsequent meetings every 6 months thereafter. Efficacy data will not be reviewed.

2 Definitions, general considerations and statistical methodology

2.1 Definitions

2.1.1 Study drug and study treatment

Study drug refers to everolimus or matching placebo.

Study treatment also refers to everolimus or matching placebo.

Note that a background AED regimen will be provided to all patients, but it will not be considered as part of the study treatment. The background AED regimen must consist of between 1 and 3 AEDs, with the exact choice and dose of AEDs left to each investigator's discretion.

2.1.2 Date of first administration of study drug

The date of first administration of ***double-blind study drug*** is defined as the first date when a non-zero dose of double-blind study drug is administered and recorded on the "Dosage administration record - Everolimus/Placebo – Core" eCRF. For the sake of simplicity, the date of first administration of study drug will also be referred to as the start of study drug.

Similarly, the date of first administration of ***everolimus in the Extension phase*** is defined as the first date when a non-zero dose of everolimus is administered and recorded on the "Dosage administration record – Everolimus – Extension" eCRF.

2.1.3 Date of last administration of study drug

The date of last administration of ***double-blind study drug*** is defined as the last date when a non-zero dose of double-blind study drug is administered and recorded on the "Dosage administration record - Everolimus/Placebo – Core" eCRF.

Similarly, the date of last administration of ***everolimus in the Extension phase*** is defined as the last date when a non-zero dose of everolimus is administered and recorded on the "Dosage administration record – Everolimus – Extension" eCRF.

2.1.4 Date of first administration of study treatment

The date of first administration of study treatment is the same as the date of first administration of study drug.

2.1.5 Date of last administration of study treatment

The date of last administration of study treatment is the same as the date of last administration of study drug.

2.1.6 Study day

The study day ***for safety assessments*** (e.g., start date of an adverse event (AE), dose interruption, neuropsychological scales, suicide scale, etc.) will be calculated using the start date of study treatment as the origin. Assessments that occur on or after the start date of study treatment will have study day calculated as (date of assessment) – (start date of study treatment) + 1. Assessments that occur before the start of study treatment will have study day calculated as (date of assessment) – (start date of study treatment). Then study day 1 for safety assessments will be the first day of study treatment, and study day -1 will be the day before the first day of study treatment.

The study day ***for all other assessments*** (e.g., seizure diary, quality of life) will be calculated using the randomization date as the origin. Assessments that occur on or after the randomization date will have study day calculated as (date of assessment) – (randomization date) + 1. Assessments that occur before the randomization date will have study day calculated as (date of assessment) – (randomization date). Then study day 1 for non-safety assessments will be the day of randomization, and study day -1 will be the day before the randomization date.

The vast majority of patients are expected to start taking study medication on the day of randomization, in which case the start date of study treatment will be the same as the randomization date.

Study day will be displayed in the data listings.

For the long-term evaluation of efficacy and safety, study day ***for all assessments*** will be calculated using the start date of everolimus as the origin. In particular, for patients originally randomized to the placebo arm, study day 1 will be the start date of everolimus in the Extension phase of the trial.

2.1.7 Baseline

Baseline is the result of an investigation describing the “true” uninfluenced state of the patient.

For ***safety evaluations*** (e.g., AEs, lab values, vital signs, neuropsychological scales, suicide questionnaire, etc.), the baseline assessment is defined as the last available assessment on or before the start date of study treatment.

For ***all other evaluations*** (e.g., seizure diary, quality of life), the baseline assessment is defined as the last available assessment on or before the randomization date. However, to reduce the risk of missing baselines, in rare cases when study treatment is started one or more days after randomization, the last available assessment prior to first treatment will be used as baseline.

If patients have no value as defined above, the baseline result will be missing.

The baseline seizure frequency is calculated from the start of the seizure diary at screening until the day before the randomization date, as defined in Section 3.9.1.1.

For the long-term evaluation of efficacy and safety, patients originally randomized to the placebo arm will have baseline values taken from the last assessment on or before the start of everolimus in the Extension phase. For example, for the Vineland-II scale, the baseline for placebo patients in the LTE Safety Set could be obtained at the end of the Core phase. Further, such patients will have baseline seizure frequency calculated over the latest 56 days of the seizure diary prior to the start date of everolimus in the Extension phase (i.e., not including the start date of everolimus).

2.1.8 On-treatment assessment/event

AE summaries will present only on-treatment assessments or events, where on-treatment means that the AE started in the following time interval (including the lower and upper limits):

- <start date of study treatment; date of last study treatment + 30 days>

Other safety summaries will present baseline and on-treatment assessments, where on-treatment means that the assessment occurred in the following time interval (including the lower and upper limits):

- <start date of study treatment + 1; date of last study treatment + 30 days>

2.1.9 Time windows

In order to summarize data over time, the following data types will be time slotted using time windows.

2.1.9.1 Neuropsychological assessments

The Vineland- II and Wechsler Nonverbal (WNV) scales will be performed at baseline, at end of Core phase (or end of treatment if patient discontinues), and then every 6 months (24 weeks) in the Extension phase plus at the end of treatment in the Extension phase.

Table 2-1 shows the defined time windows for the neuropsychological assessments.

Table 2-1 Time windows for neuropsychological assessments

Time Window	Planned Visit Timing	Time Window Definition
Baseline	On or before Study Day 1*	≤ Study Day 1
Week 9	Study Day 64	Study Days 2 – 94
Week 18	Study Day 127	Study Days 95 – 211
Week 42	Study Day 295	Study Days 212 – 379
Week 66	Study Day 463	Study Days 380 – 547
Every 24 weeks thereafter		
Week =66+24×k, (where k=1,2,3,…)	Study Day = (66+24×k) ×7+1	Study Days (66+24×k) ×7+1 – (12×7-1) to (66+24×k) ×7+1 + (12×7)

* Study Day 1 = start date of study treatment (or for long-term evaluation, study day 1 = start date of everolimus)

Excluding the baseline, for which the last value is used, if more than one assessment is done within the same time window, the assessment performed closest to the target date will be used; if two or more assessments are equidistant from the planned date, then the mean values will be used.

2.1.9.2 Everolimus pharmacokinetics

Blood samples will be collected for everolimus pharmacokinetic (PK) analysis at weeks 1, 3, 5, 10, 14 and 18 during the Core phase, and at weeks 19, 22, 26, and 30, and every 12 weeks thereafter during the Extension phase. Additional everolimus PK samples will be collected approximately 2 weeks, but not less than 5 days after any change in the dose of study medication, or approximately 2 weeks, but not less than 5 days after after any change in use of concomitant CYP3A4/P-glycoprotein (PgP) inhibitors or inducers.

Table 2-2 shows the defined time windows for everolimus PK.

Table 2-2 Time windows for everolimus PK

Time Window	Planned Visit Timing	Time Window Definition
Week 1	Study Day 8*	Study Days 2 – 15
Week 3	Study Day 22	Study Days 16 – 29
Week 5	Study Day 36	Study Days 30 – 53
Week 10	Study Day 71	Study Days 54 – 85
Week 14	Study Day 99	Study Days 86 – 113
Week 18	Study Day 127	Study Days 114 – 130
Week 19	Study Day 134	Study Days 131 – 145
Week 22	Study Day 155	Study Days 146 – 169
Week 26	Study Day 183	Study Days 170 – 197
Week 30	Study Day 211	Study Days 198 – 251
Week 42	Study Day 295	Study Days 252 – 337
Every 12 weeks thereafter		
Week =42+12×k, (where k=1,2,3,…)	Study Day = (42+12×k)×7 +1	Study Days (42+12×k)×7 +1 – (6×7 - 1) to (42+12×k)×7 +1 + (6×7)

* Study Day 1 = start date of everolimus

If more than one assessment is done within the same time window, the mean value per time window will be used.

For the long-term evaluation of everolimus, placebo patients who enter the Extension phase will be included. As the planned schedule of blood draws for everolimus PK is different in the Extension phase compared with the Core phase, the time windows in Table 2-3 will be used in some analyses for such patients.

Table 2-3 Time windows for everolimus PK – patients randomized to the placebo arm

Time Window	Planned Visit Timing	Time Window Definition
Week 1	Study Day 8*	Study Days 2 – 18
Week 4	Study Day 29	Study Days 19 – 43
Week 8	Study Day 57	Study Days 44 – 71
Week 12	Study Day 85	Study Days 72 – 127

Time Window	Planned Visit Timing	Time Window Definition
Week 24	Study Day 169	Study Days 128 – 211
Every 12 weeks thereafter		
Week =24+12×k, (where k=1,2,3,…)	Study Day = (24+12×k)×7 +1	Study Days (24+12×k)×7 +1 – (6×7 - 1) to (24+12×k)×7 +1 + (6×7)
* Study Day 1 = start date of everolimus		

2.1.9.3 AED pharmacokinetics

Blood samples will be collected from all patients for AED pharmacokinetic (PK) analysis at week -8 (screening), on study day 1 (baseline), and at weeks 1 and 3 during the Core phase. Only samples from patients randomized to one of the two everolimus arms will be analyzed.

Table 2-4 shows the defined time windows for AED PK.

Table 2-4 Time windows for AED PK

Time Window	Planned Visit Timing	Time Window Definition
Week -8	Study Day -56*	Study Days -70 to -8
Baseline	Study Day 1	Study Days -7 to 1
Week 1	Study Day 8	Study Days 2 to 15
Week 3	Study Day 22	Study Days 16 to 43
* Study Day 1 = start date of everolimus		

If more than one assessment is done within the same time window, the mean value per time window will be used.

2.1.9.4 Columbia Suicide Severity Rating Scale

The electronic Columbia Suicide Severity Rating Scale (eC-SSRS) should be performed at every visit (except the PK visits at weeks 1, 3, 5 and 19, and the end of study visit).

Table 2-5 shows the defined time windows for eC-SSRS.

Table 2-5 Time windows for eC-SSRS suicide scale

Time Window	Planned Visit Timing	Time Window Definition
Screening	Study Day -56*	Study Days -70 to -8
Baseline	Study Day 1	Study Days -7 to 1
Week 2	Study Day 15	Study Days 2 – 22
Week 4	Study Day 29	Study Days 23 – 36
Week 6	Study Day 43	Study Days 37 – 57
Week 10	Study Day 71	Study Days 58 – 85
Week 14	Study Day 99	Study Days 86 – 113
Week 18	Study Day 127	Study Days 114 – 141
Week 22	Study Day 155	Study Days 142 – 169
Week 26	Study Day 183	Study Days 170 – 197
Week 30	Study Day 211	Study Days 198 – 251
Week 42	Study Day 295	Study Days 252 – 337
Every 12 weeks thereafter		
Week =42+12×k, (where k=1,2,3,…)	Study Day = (42+12×k)×7 +1	Study Days (42+12*k)*7+1-(6*7-1) to (42+12×k)×7 +1 + 6×7

* Study Day 1 = start date of everolimus

These time windows will only be used to summarize compliance, and so no rules are required for cases when there is more than one assessment in the same time window.

2.1.9.5 Growth data

Growth data (height and weight) will be collected at every visit (except the PK visits at weeks 1, 3, 5 and 19, and the end of study visit). There is also an attempt to retrospectively collect growth data prior to screening, and so pre-baseline time windows will be required. Note that for the long-term evaluation, for patients randomized to the placebo arm, pre-baseline also includes the double-blind Core phase of the trial, prior to the start of everolimus in the Extension phase.

Table 2-6 summarizes the time windows for growth data, where windows are centered at every 24 weeks before and after start of everolimus. Although height and weight are collected more frequently than every 24 weeks (post-randomization), this choice of time window length was made to reflect the degree of accuracy in the reference values (every 6 months) that will be used in the calculation of summary variables of growth (see Section 3.11.1).

Table 2-6 Time windows for growth data

Time Window	Planned Visit Timing	Time Window Definition
Week -72	Study Day -504	Study Days -588 to -421
Week -48	Study Day -336	Study Days -420 to -253
Week -24	Study Day -168	Study Days -252 to -85
Baseline	Study Day 1	Study Days -42 to 1
Week 24	Study Day 169	Study Days 86 to 253
Week 48	Study Day 337	Study Days 254 to 421
Week 72	Study Day 505	Study Days 422 to 589
Every 24 weeks thereafter		
Week =72+24×k, (where k=1,2,3,…)	Study Day = (72+24×k)×7 +1	Study Days (72+24×k)×7 +1 – (12×7 -1) to (72+24×k)×7 +1 + (12×7)

Study Day 1 = start date of everolimus

If more than one assessment is done within the Baseline time window, the assessment closest to treatment day 1 will be used; if two or more assessments are equidistant from treatment day 1, then the mean value will be used. For all other time windows, the assessment closest to the planned assessment date will be used; if two or more assessments are equidistant from the planned date, then the mean value will be used.

2.1.10 Definitions of analysis sets

Full Analysis Set

The Full Analysis Set (FAS) comprises all patients to whom study treatment has been assigned by randomization. Following the intent-to-treat principle, patients will be analyzed according to the treatment arm and stratification factor that they were assigned to at randomization. The FAS is the primary efficacy population.

Safety Set

The Safety Set includes all patients who received at least one dose of study medication in the Core phase, and had at least one post-baseline safety assessment in the Core phase (where the statement that a patient had no AE (on the AE eCRF) constitutes a safety assessment).

There are three treatment groups in the study: everolimus titrated to a trough concentration of 3 to 7 ng/ml, everolimus titrated to a trough concentration of 9 to 15 ng/ml, and placebo. The usual approach for the Safety Set would be to analyze patients according to study treatment received. However, that is not possible because the two everolimus arms have the same treatment received (i.e., everolimus), and are only differentiated by differing targeted trough ranges. Therefore, patients randomized to one of the two everolimus arms and who received at least one dose of everolimus, will be analyzed in the Safety Set according to the everolimus treatment arm assigned at randomization. Similarly, patients randomized to placebo and who received at least one dose of placebo, will be analyzed in the placebo arm of the Safety Set.

That is, only patients who received the wrong treatment regimen throughout their entire time in the Core phase will be reassigned to a different treatment arm in the Safety Set compared with the FAS, as follows:

- Patients randomized to either of the two everolimus arms and who received only placebo tablets during the Core phase will be reassigned to the placebo arm in the Safety Set. This could happen for example if the first kit of study medication contained only placebo tablets (i.e., given in error), and the patient took at least one tablet and then subsequently discontinued before taking any everolimus tablets.
- Patients randomized to the placebo arm and who received only everolimus tablets during the Core phase will be reassigned to the everolimus 3 to 7 ng/ml arm in the Safety Set. Again this would require a dispensing error in the first kit of study medication, followed by early discontinuation. Note that the reassignment is to the lower everolimus targeted trough arm, chosen because such an error in starting dose would be recognized via IRT before any dose increase could be made.

Note that there will be no attempt to reassign patients from one everolimus arm to the other everolimus arm based on observed everolimus concentrations during the Core phase. For example, a patient randomized to the everolimus 9 to 15 ng/ml arm and whose observed everolimus trough concentrations ranged from 3 to 7 ng/ml during the entire Core phase would still be considered to have treatment received = everolimus 9 to 15 ng/ml. This is because increases in the dose of everolimus would have been recommended during the Core phase in order to reach the targeted trough range of 9 to 15 ng/ml, and any associated toxicity should therefore be attributed to the 9 to 15 ng/ml arm and not the 3 to 7 ng/ml arm.

Per Protocol Set

The Per-Protocol Set (PPS) consists of a subset of patients in the FAS who are compliant to the requirements of the protocol, who are evaluable for efficacy and who have completed a minimum exposure requirement. However, if a patient discontinued for lack of efficacy, AE or death before the minimum exposure requirement could be met, or before he/she could become evaluable for efficacy, that patient will still be included in the Per Protocol Set.

Patients will be evaluable for efficacy if their seizure diary continues until at least day 70 of the Core phase, i.e., at least the first 4 weeks of the maintenance period of the Core phase.

The minimum exposure requirement is defined as receiving study treatment on ≥ 50% of days during the first 10 weeks of the Core phase.

Major protocol deviations excluding patients from the PPS are described in Section 2.3.

Long-Term Evaluation Efficacy Set

The Long-Term Evaluation Efficacy Set (LTE Efficacy Set) is intended to capture all data on everolimus in the trial, from both the Core phase and the Extension phase. It consists of all patients who receive at least one dose of everolimus in the trial and have at least one post-baseline efficacy measure (i.e., at least one day on which the number and types of seizures are known and documented in the "Investigator's Seizure Classification" eCRF). Patients will be analyzed as a single everolimus group as well as separately according to their randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3 to 7 ng/ml and 9 to 15 ng/ml, or placebo). This analysis set will be used for long-term efficacy analyses of everolimus.

Before protocol amendment 2, patients from the two everolimus groups were to continue in the Extension phase with the same targeted trough range from Core phase, while placebo patients where switched to receive everolimus during the Extension phase with a 3-15 ng/ml targeted trough. Protocol amendment 2 introduced the possibility for investigators to make the dose adjustments in the Extension phase based on the C_{min} value. If investigators chose not to modify dosing, the dose was determined by the IRT to maintain the everolimus trough concentration range between 3-15 ng/ml for all patients. In order to avoid compromising the blinded nature of the study (including the investigator, patient and sponsor) in the earlier Core phase, all patients were be transitioned towards a common targeted trough range of 6 to 10 ng/ml before any C_{min} value is shared with the investigator. For this reason, for analyses of long-term evaluation data from the Core and Extension phase, patients will be analyzed as a single everolimus group ("All patients") as well as separately according to the randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3 to 7 ng/ml and 9 to 15 ng/ml and placebo). Note that "placebo" arm will only contain data from the Extension phase when patients received everolimus.

Long-Term Evaluation Safety Set

The Long-Term Evaluation Safety Set (LTE Safety Set) is intended to capture all data on everolimus in the trial, from both the Core phase and the Extension phase. It consists of all patients who receive at least one dose of everolimus in the trial and have at least one post-baseline safety measure. Patients will be analyzed as a single everolimus group as well as separately according to their randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3 to 7 ng/ml and 9 to 15 ng/ml, or placebo). This analysis set will be used for long-term safety and pharmacokinetic analyses of everolimus.

For the same reason as described above for the LTE Efficacy Set, patients will be analyzed as a single everolimus group ("All patients") as well as separately according to their randomized treatment arm in the Core phase (i.e., everolimus trough ranges of 3 to 7 ng/ml and 9 to 15

ng/ml and placebo). Note that "placebo" arm will only contain data from the Extension phase when patients received everolimus.

Pharmacokinetic Set

There will be no formal pharmacokinetic set. The pharmacokinetic analyses will be performed in the Safety Set and the LTE Safety Set using all available PK samples Set or PK Sensitivity Sample Set, as defined below (section 3.13.1.1).

2.2 General considerations

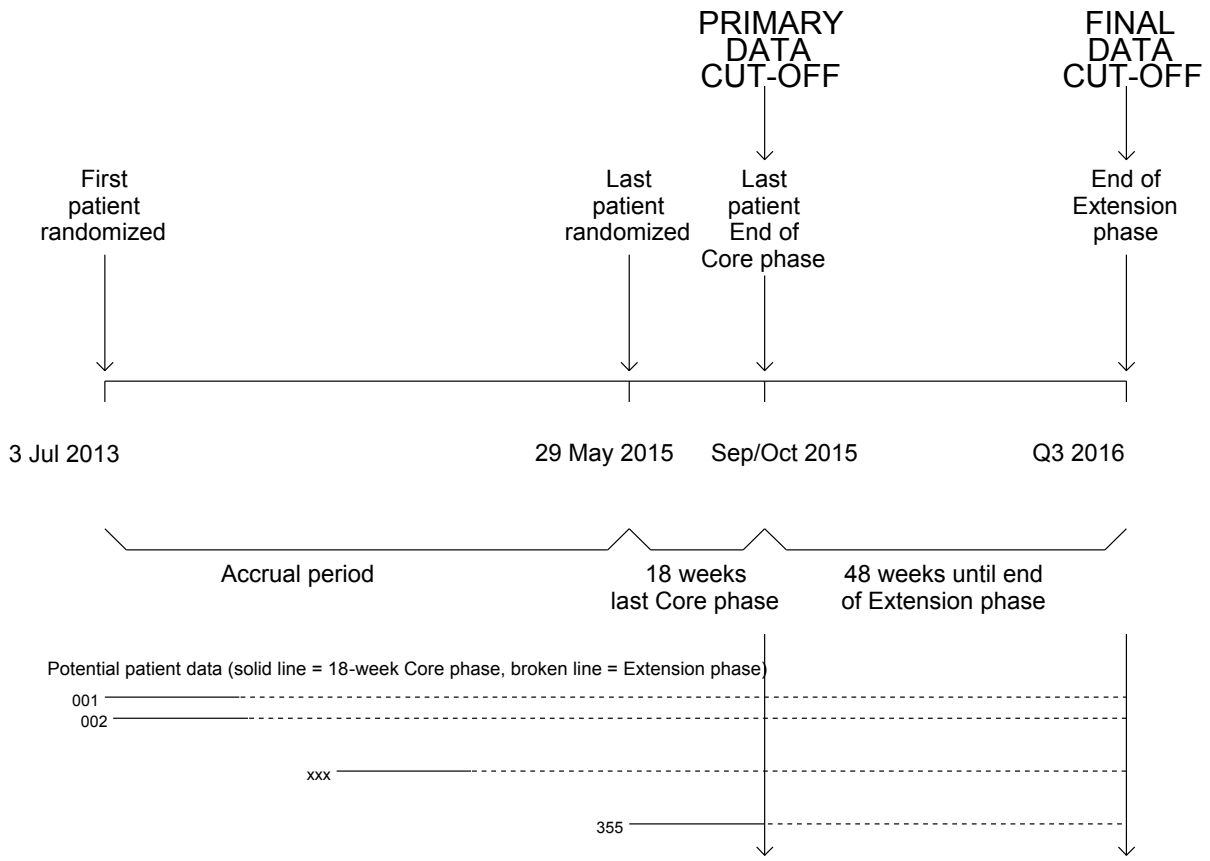
2.2.1 Data included in the analysis

It is planned that the data from all centers participating in the trial will be combined, so that an adequate number of patients are available for analysis. Novartis and/or a designated CRO will perform all the analyses.

The primary data cut-off date, for the ***primary analysis***, will be when all patients have completed the Core phase or have discontinued early (see Figure 2-1 below). At that time, all data from the Core phase will be available as will a part of the Extension phase data. The analyses will thus cover both periods, although the main focus will be on inferential and descriptive comparisons of each everolimus arm versus placebo on efficacy and safety in the Core phase. In addition, descriptive analyses for the long-term evaluation of everolimus will be presented based on data from the Core phase and the available data from the Extension phase. All planned analyses for the Core phase will be produced at that time. Only a subset of the analyses for the long-term safety and efficacy evaluation of everolimus will be produced. All analyses for the long-term evaluation will be produced after the final cut-off at the end of extension phase.

Figure 2-1 shows that the first patient was randomized on July 3, 2013, and that accrual of the planned 355 patients was completed on May 29, 2015. Then the data cut-off date for the primary analysis will be end of September/beginning of October 2015, assuming the last patients complete the Core phase. The figure also shows the extent of Extension phase data at the primary cut-off date, with the first randomized patients potentially having more than 12 months follow-up during the Extension phase.

Figure 2-1 Example timing of data cut-off dates



There will be a final data cut-off date at the end of the Extension phase, for the ***final analysis***, which is planned 48 weeks after the last patient completes the Core phase (except if the primary analysis fails to show a benefit for everolimus over placebo, in which case the trial will be terminated). The analyses at the final data cut-off will complete the long-term evaluation of everolimus, using all data on everolimus from both the Core and Extension phases of the study. As data from the Core phase will remain unchanged since the first cut-off date, no new analyses comparing the everolimus arms to placebo will be undertaken at the final analysis.

No other analyses between the primary analysis and the final analysis are currently planned, but may be performed if needed.

The tables, listings and figures to be produced at each analysis (primary and final) will be identified in RAP Module 7. All data collected in the database will be used for the statistical analysis.

Only data with an assessment date or event start date on or before the data cut-off date will be included in the analysis. For example, if the data cut-off date is 15-Jun-2014 then an AE starting on 13-Jun-2014 will be reported, whereas an AE with start date on 17-Jun-2014 will not be reported.

All events with a start date on or before the cut-off date and with an end date after the cut-off date will be reported as 'continuing at the cut-off date'. The same rule will apply to events starting on or before the cut-off date and without a documented end date. This approach applies in particular to AE and concomitant medication reports. For these events, the cut-off date will not be imputed and therefore will not appear in the listings.

If it is required to impute an end date in order to perform a specific analysis, the imputed date will be displayed and flagged in the listings. For example, if there is a record in the Dosage Administration Record (DAR) panel with missing end date (or end date after the cut-off date), the cut-off date needs to be imputed as the end date in order to enable treatment exposure duration and dose intensity to be calculated. Details of imputation of partial dates will be provided in the programming specifications (RAP Module 8).

2.2.2 Patient Classification

Patients are excluded from the analysis sets defined above based on protocol deviations and/or specific patient classification rules, as defined in Table 2-7. For example, a patient will be excluded from the Safety Set if they meet either of the conditions (i) no post-baseline safety assessment or (ii) no dose of study medication (as programmed in derived dataset *noviopto*).

Table 2-7 Patient classification rules

Analysis set	Protocol deviation severity codes leading to exclusion	Additional patient classification rules leading to exclusion*
Full Analysis Set	8	Patient not randomized.
Safety Set	8, 5	(i) Patient with no post-baseline safety assessment. (ii) Patient with no dose of study medication, i.e., no non-zero dose in the "Dosage administration record - Everolimus/Placebo – Core" eCRF
Per Protocol Set	8, 1	(i) Patient not randomized. (ii) Not evaluable for efficacy (seizure diary discontinued during first 10 weeks of Core phase). ** (iii) Insufficient treatment exposure (received study treatment on < 50% of days during first 10 weeks of Core phase). ** (iv) Received the wrong treatment regimen throughout the entire duration of the Core***

Severity code: 1=exclude from per protocol analysis, 5=exclude from all safety analyses, 8=exclude from all analyses (to be used only if patient has not signed informed consent prior to start of study)
* Classification rule will be determined programmatically
** However, if a patient discontinued for lack of efficacy, AE or death before the minimum exposure requirement could be met, or before he/she could become evaluable for efficacy, that patient will still be included in the Per Protocol Set.
*** This major deviation will be determined programmatically after the unblinded information of IRT is available, in order to determine which treatment was received by the patients. Since it cannot be known before unblinding, this information will not be found in dataset of protocol deviations (Viopto). This deviation, if any, will have to be added to the derived dataset a_viopto in order to be reported in the summary of protocol deviations.

All protocol deviations and patient classification rules will be finalized before database lock.

2.3 Major protocol deviations

The following protocol deviations are considered as major deviations and constitute reasons for excluding patients from the per-protocol set:

- Absence of diagnosis of TSC or partial-onset epilepsy.
- Previously failed < 2 AEDs given individually or combined.
- Prior therapy with systemic or topical mTOR inhibitors within the exclusion criteria time windows.
- Baseline seizure diary with < 16 ESC-approved partial-onset seizures.
- Baseline seizure diary containing a seizure-free period of ≥ 21 consecutive days.
- Baseline seizure diary < 49 days in duration.
- Did not receive 1-3 AEDs at same dose from 4 weeks prior to screening visit to baseline visit
- Change in dose for more than 7 days or in number of concomitant AEDs during Core Phase or interruption of AED > 7 days
- Missing seizure information on > 50% of days in the Core phase or the Baseline phase.
- Received the wrong treatment regimen throughout their entire time in the Core*, defined as:
  - Patients randomized to either of the two everolimus arms and who received only placebo tablets during the Core phase.
  - Patients randomized to the placebo arm and who received only everolimus tablets during the Core phase.

* This deviation will be determined programmatically, post database lock and unblinding. If an occurrence is found, it will be added to the a_viopto dataset in order to be reported in the deviation summary tables.

Other protocol deviations will also be identified, summarized and listed. However, patients will not be excluded from the Per-Protocol set based on these other deviations.

2.4 Concomitant medications with specific impact on the analysis

Each of the following categories of medication or non-drug therapy will be identified based on a clinical review of medication and non-drug therapies listings. Details on how each list should be prepared for review, be updated and used to produce summary tables and listings will be provided in the programming specification (RAP Module 8).

2.4.1 Background AED regimen

Patients are required to be on a stable background AED regimen consisting of between 1 and 3 different AEDs, starting from 4 weeks prior to screening and continuing until the end of the Core phase. The background AED regimen will be recorded on the DAR – Antiepileptic Therapy eCRF page.

2.4.2 Rescue medications

Rescue medications are allowed when patients experience transient increases in seizure activity despite best medical management. Rescue medications are essentially benzodiazepines (e.g., buccal midazolam, rectal diazepam), and should be prescribed according to the local institution's practice. Rescue medications will be recorded on the rescue medications eCRF page.

2.4.3 Inducers and inhibitors of CYP3A4 and/or PgP

Many patients will be taking concomitant CYP3A4 enzyme inducing drugs as part of their background regimen of antiepileptic drugs. Therefore, co-administration of strong inhibitors of CYP3A4 or P-glycoprotein (PgP) should be avoided and moderate inhibitors used with caution. These substances are listed in the Investigator's Brochure and referenced in the study protocol.

However, some patients may take these substances during the trial so these concomitant medications will be selected via programming and will be tabulated and listed. An everolimus project level document lists all such known substances, and ongoing review by the Clinical Pharmacologist will be performed as follows:

- Substances will be identified as inhibitors, inducers, or substrates
- Inhibitors of CYP3A4 will be also classified as strong, moderate, weak, or other
- Inhibitors of PgP will be classified as yes or no
- Inducers of CYP3A4 will be classified as yes or no
- Inducers of PgP will be classified as yes or no
- Substrates of CYP3A4 will be classified as yes or no
- Substrates of PgP will be classified as yes or no

Substances that have not been reviewed and those which are no longer available in the current version of the WHO dictionary will be flagged.

Review listings will be produced for this study but will always feed back into the project level list. This list is considered applicable at the everolimus project level and is to be used for all clinical indications.

A corresponding list for programming purposes will be stored in the Novartis global programming system (i.e., in GPS2, in the compound level CRAD001/util folder).

2.5 Seizure data

The primary efficacy endpoint of the study is change from baseline in frequency of partial-onset seizures. This section gives more detailed information on the process of how seizure events are collected, reviewed and documented, and on which seizure events are included in counts of partial-onset seizures for the primary endpoint.

2.5.1 Investigator/Patient interview at the screening visit

Seizures are recorded in patient seizure diaries on an ongoing basis throughout the trial. During the screening visit, patients and/or caregivers will be asked by the investigator to

describe the patient's seizures in their own words. There may be certain events that are considered to be an epileptic seizure by the patient, but that are considered to be behavioral events by the investigator. In any case, the investigator is required to use his/her expert judgment to identify events that in his/her opinion have at least a 50% probability of being an epileptic seizure, and to request the patient or caregiver to only record such events in their seizure diary. These events will be defined on the first page of the patient seizure diary using a simple coding scheme like A, B, C, etc. The patient then embarks on the 8-week baseline phase, and records each occurrence of A, B, C, etc. in the patient seizure diary as they occur.

2.5.2 Independent review by the Epilepsy Study Consortium

An independent review of the seizure types being recorded by the patient will be performed by the Epilepsy Study Consortium (ESC) in conjunction with the investigator. The ESC is an independent group of scientific investigators from academic medical research, dedicated to accelerating the development of new therapies in epilepsy to improve patient care.

Within 48 hours of the screening visit, the investigator will complete a Diagnostic Review Form (DRF) which documents what information was used to determine the patient's diagnosis and provides the seizure history, with supporting materials if necessary (e.g., EEGs). In addition, the investigator completes a Seizure Identification Form (SIF) for each seizure, answering questions about the event and providing a seizure code from a list that is reproduced below in Table 2-8.

Table 2-8 Seizure codes

Seizure type		Protocol definition of partial-onset*
FOCAL SEIZURE		
IA1	Motor or autonomic components, but without impairment of consciousness or awareness	✓
IA2a	Subjective sensory or psychic phenomena; without impairment of consciousness or awareness; with ictal EEG confirmation	✓
IA2b	Subjective sensory or psychic phenomena; without impairment of consciousness or awareness; without ictal EEG confirmation	×
IB1	Complex partial seizures, predominantly stare	✓
IB2	Complex partial seizures, predominantly stare and facial automatisms	✓
IB3	Complex partial seizures (Not Otherwise Specified)	✓
SECONDARILY GENERALIZED		
IC1	Convulsive seizure: Myoclonic features	✓
IC2	Convulsive seizure: Clonic features	✓
IC3	Convulsive seizure: Tonic features	✓
IC4	Convulsive seizure: Tonic-clonic features	✓
IC5	Convulsive seizure: Atonic features	✓
IC6	Convulsive seizure: NOS	✓
GENERALIZED ONSET SEIZURES		
IIA1	Convulsive seizure: Typical Absence seizures	×
IIA2	Convulsive seizure: Atypical Absence seizures	×
IIB	Convulsive seizure: Myoclonic seizures	×
IIC	Convulsive seizure: Clonic seizures	×
IID	Convulsive seizure: Tonic seizures	×

Seizure type		**Protocol definition of partial-onset***
IIE	Convulsive seizure: Tonic-clonic seizures	×
IIF	Convulsive seizure: Atonic seizures	×
IIG	Convulsive seizure: NOS	×

*: Protocol definition of partial-onset seizures: seizures with motor signs that have not been documented to be generalized onset seizures on EEG or sensory seizures confirmed to be partial-onset by ictal EEG – see Section 2.5.4 below.

*: ✓ indicates seizure type qualifying for protocol definition of partial-onset; × seizure type excluded from the protocol definition of partial-onset

The SIF also solicits the subjective judgment of the investigator as to whether the seizure is Probable (greater than 80% likelihood of being an epileptic seizure) or Questionable (50-80% chance). In fact the aim in this study is to only record events in the eCRF that have a high probability of being seizures, and one of the main tasks of the ESC review is to ensure sites are consistently identifying such seizures.

The DRF and SIF forms are sent to the ESC for their review, and this process will be followed for every patient in the study. With the benefit of having this overview across all investigators, and in the interests of making seizure classification as consistent as possible across sites, the ESC can make a recommendation to an investigator to change a particular seizure classification, although the final decision rests with the investigator. For example, some seizures that were initially considered Probable by the investigator may end up not getting ESC approval, whereas other seizures rated as Questionable by the investigator may get approved.

ESC findings will be communicated to the investigator via a Feedback Form. Questions or requests for additional information made by the ESC will be documented using the Feedback Form. If corrections, clarifications or additional information are required, the site is instructed to respond on the Feedback Form and depending on the response, may be asked to revise and resubmit the DRF and/or SIF. The documents are then to be resubmitted to the ESC for a second review.

Once agreement has been reached between the ESC and the investigator, an Approval Form will be sent, containing a list of events to be counted as seizures (including the agreed upon seizure code as per Table 2-8), and a list of events not to be counted as seizures in the eCRF.

If a new seizure type occurs during the trial, the investigator will complete a new SIF and submit it to the ESC for review. If new information about a previous seizure type becomes available, the investigator should revise the DRF and/or SIF and resubmit to the ESC. Then a seizure that was originally not approved, could on the basis of new information, be reclassified as approved, or vice versa. In either case, the investigator would need to review all previous Investigator Seizure Classification eCRFs in order to make the necessary corrections, starting from the beginning of the Baseline phase.

If the investigator and the reviewer do not agree on the classification/diagnosis, the forms will be sent to a second ESC reviewer as a quality control step. In the event that the investigator and the ESC do not resolve their difference, the decision of the investigator is final.

2.5.3 Transcription of patient seizure diary to eCRF

When entering the patient seizure diary into the Investigator Seizure Classification eCRF, the investigator should only include events counted as seizures on the Approval Form. However, patients should continue to track all events, as originally instructed by the investigator, in their seizure diaries (i.e., events counted as seizures as well as events not counted as seizures).

2.5.4 Seizures counted towards partial-onset seizure frequency

The study protocol gave a specific definition of partial-onset seizures that will be counted in this trial: seizures with motor signs that have not been documented to be primary generalized seizures on EEG. The study protocol amendment 2 further extended this definition by including also the sensory seizures (without motor sign) confirmed to be partial-onset by ictal EEG. Considering the list of seizure codes provided in Table 2-8, this excludes code IA2b because such seizures do not include any motor component and were not confirmed by EEG, and also codes IIA1 through IIG which are (primarily) generalized onset seizures. Therefore, the seizure codes that count towards the primary endpoint are: IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 and IC6.

2.6 Stratification

2.6.1 Stratified randomization

The randomization will use stratified random blocks with a single stratification factor of patient age at randomization split into 4 distinct groups as follows: 1 to <6 years; (ii) 6 to <12 years; (iii) 12 to <18 years; and (iv) ≥ 18 years.

2.6.2 Stratified analysis

A stratified analysis method will be used to test for treatment group differences, since the factor on which the randomization has been stratified should be accounted for in the analysis (ICH guideline E5).

The stratified CMH test and the stratified rank ANCOVA model will be the primary analyses for the EMA and FDA primary endpoints respectively (see Section 3.9.1).

3 Statistical methods used in reporting

3.1 General presentation of descriptive summaries

Qualitative data (e.g., gender, race, etc.) will be summarized by means of contingency tables by treatment group; a missing category will be included as applicable. Percentages will be calculated using the number of patients in the relevant population or subgroup as the denominator.

Quantitative data (e.g., age, weight, etc.) will be summarized by appropriate descriptive statistics (i.e., mean, standard deviation, median, minimum, maximum) by treatment group.

3.2 Enrollment status

Number of patients screened will be summarized by country and center. Number of patients randomized will be summarized by country, center and treatment group. This information will also be presented separately for the FAS, Safety Set, Per-Protocol Set and the LTE Efficacy and Safety sets.

3.3 Background and demographic characteristics

The FAS and the LTE Efficacy Set will be used for all baseline and demographic summaries and listings.

3.3.1 Basic demographic and background data

All demographic and background data will be listed. A summary table will include age, sex, weight, height, body mass index (BMI), body surface area (BSA), race, ethnicity and puberty stage at baseline.

Qualitative data will be summarized by means of contingency tables, and quantitative data will be summarized by appropriate descriptive statistics (mean, standard deviation, median, minimum, and maximum).

3.3.2 Stratification factors

The randomization stratification factor actually used to randomize patients (contained in the IVR dataset) will be tabulated and listed against patient age subgroups calculated from date of birth and randomization date as collected on the eCRFs.

3.3.3 Protocol eligibility criteria

Protocol eligibility criteria as per eCRFs at randomization will be summarized and listed.

3.3.4 Diagnosis and extent of TSC disease

Diagnosis of TSC will be summarized according to the criteria listed in Table 3-1 below.

Table 3-1 Diagnostic Criteria for Tuberous Sclerosis Complex (TSC)

Major Features
1. Facial angiofibromas or forehead plaque
2. Non-traumatic ungual or periungual fibroma
3. Hypomelanotic macules (three or more)
4. Shagreen patch (connective tissue nevus)
5. Multiple retinal nodular hamartomas
6. Cortical tuber [a]
7. Subependymal nodule
8. Subependymal giant cell astrocytoma
9. Cardiac rhabdomyoma, single or multiple
10. Lymphangioleiomyomatosis [b]
11. Renal angiomyolipoma [b]
Minor Features

1. Multiple, randomly distributed pits in dental enamel
2. Hamartomatous rectal polyps [c]
3. Bone cysts [d]
4. Cerebral white matter radial migration lines [a, d]
5. Gingival fibromas
6. Non-renal hamartoma [c]
7. Retinal achromic patch
8. 'Confetti' skin lesions
9. Multiple renal cysts [c]

Definite Tuberous Sclerosis Complex:
Two Major Features or one Major Feature plus two Minor Features.

a. The co-occurrence of cerebral cortical dysplasia and cerebral white matter radial migration lines should be considered as one major feature of TSC.

b. In patients with both lymphangioleiomyomatosis and renal angiomyolipoma, another feature of TSC must be identified before a definite diagnosis is assigned.

c. Histologic confirmation of these features is suggested.

d. Radiographic confirmation of these features is sufficient.

Time since initial diagnosis of TSC-related conditions or symptoms will be summarized in years (defined as 365.25 days), and will be measured until the time of baseline (i.e., start date of study treatment).

3.3.5 Diagnosis and extent of epilepsy

Summary statistics will be tabulated for diagnosis and extent of epilepsy. These analyses will include seizure history (status epilepticus, simple partial seizure, complex partial seizure, secondarily generalized seizure, generalized onset seizure, other) and time since initial diagnosis or partial-onset seizures.

3.3.6 Medical history

Medical history and ongoing medical conditions, including TSC-related conditions and symptoms, will be summarized and listed by treatment group. Separate summaries will be presented for historical medical conditions (marked as "not active" on the Medical History eCRF) and for ongoing medical conditions (marked as "active"). In addition, pre-treatment AEs will be included in the summary of ongoing medical conditions, defined as any AE starting between the time of signing informed consent and the day before starting study treatment.

The summaries will be presented by primary system organ class and preferred term. Medical history/current medical conditions are coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology.

3.3.7 Prior anti-TSC therapy

Prior anti-TSC therapy will be listed, and the number and percentage of patients recording any prior anti-TSC medications will be summarized by treatment group.

Prior anti-TSC medications will be summarized by Anatomic Therapeutic Chemical (ATC) class, preferred term and treatment group.

3.3.8 Prior antiepileptic therapy

The number and percentage of patients with prior epilepsy surgery, prior vagal nerve stimulator (VNS) and prior ketogenic diet will be summarized by treatment group.

The number of prior AEDs within 1 year prior to screening, 3 years prior to screening and over the patient's lifetime will be summarized by treatment group.

Prior AEDs will also be summarized by therapy type, ATC class, and preferred term, within 1 year prior to screening, 3 years prior to screening, as well as throughout the entire patient's life.

3.3.9 TSC genetic testing

Proportions of patients with TSC1 and TSC2 gene mutations will be presented per treatment arm.

3.3.10 Pregnancy history, menstrual history and monitoring

This data is collected to provide background and additional information for observed cases of amenorrhea (see Section 3.10.1.5), and will be listed.

3.3.11 Other

All other data collected at baseline, including source of subject referral, child bearing potential, pregnancy test results, and biomarker informed consent, will be listed.

3.4 Protocol deviation summaries

The number and percentage of patients in the FAS with any protocol deviation will be tabulated by the deviation category (as specified in the Validation and Planning (VAP) documents) and by treatment group. Protocol deviations will also be summarized by center.

Protocol deviations leading to exclusion from the Per Protocol Set will be tabulated separately by treatment group. All protocol deviations will be listed, with deviations leading to exclusion from the Per Protocol Set being flagged.

3.5 Groupings for Analysis

The number and percentage of patients in each analysis population will be summarized by treatment group. The distribution of patients at screening and in the FAS, Per Protocol Set, Safety Set and the LTE Efficacy and Safety Sets will also be summarized by country, center and treatment group.

3.6 Patient disposition

The FAS will be used for the patient disposition summary tables and listings which will be summarized separately for the Core phase and the Extension phase.

3.6.1 Core phase

Information from the End of Treatment (Core phase) eCRF will provide a summary by treatment arm showing:

1. Number (%) of patients who completed the Core phase.
2. Number (%) of patients who discontinued double-blind study treatment:
   a. During titration period.
   b. During maintenance period.
   c. Overall.
3. Reasons for study treatment discontinuation.

3.6.2 Extension phase

Information from the End of Treatment (Extension phase) eCRF will provide a summary by treatment arm, and using the number of patients entering the Extension phase as denominator, showing:

1. Number (%) of patients who completed the Extension phase.
2. Number (%) of patients who discontinued from the Extension phase.
3. Reasons for discontinuation from the Extension phase.

3.7 Study treatment

Duration of study treatment exposure, cumulative dose and dose intensity (DI) will be summarized by treatment group in the Core phase using the Safety Set, and separately for the LTE Safety Set using all data on everolimus from the Core and Extension phases.

In addition, the duration of study treatment exposure will be categorized into time intervals, and frequency counts and percentages will be presented for the number of patients in each interval. The number of patients with dose reductions or interruptions, and the reasons, will be summarized by treatment.

Dose modification is planned for the study, with all dose modifications recommended by IRT based on everolimus trough concentrations and the targeted trough range (3 to 7 ng/ml or 9 to 15 ng/ml depending on the treatment arm assigned at randomization). In order to maintain the blind, patients in the placebo arm may receive dose changes, and patients in the everolimus 3 to 7 ng/ml arm may receive placebo tablets as dummy dose increases.

Dose of study treatment will be defined in units of mg/m^2, calculated as the actual dose received in mg divided by the most recent determination of BSA obtained on or before the date of dosing. BSA is calculated using the Dubois and Dubois (1916) formula:

$$BSA = (W^{0.425} \times H^{0.725}) \times 0.007184,$$

where W denotes weight in kilograms and H denotes height in centimeters.

All patients are started at a dose of between 3 and 9 mg/m^2 depending on three categories of age at randomization and whether the patient is receiving a CYP3A4/PgP inducer (see [Table 6-2 of the study protocol] for details). Study treatment is given as 2 mg dispersible tablets, and the number of tablets to be administered should be calculated from the dose per mg/m^2 and the patient's BSA, and using the rounding rules described in [Section 6.2.3 of the study protocol]. Doses at subsequent visits can be increased in order to reach the targeted everolimus trough concentration range, but only by one whole 2 mg tablet (or two 2 mg tablets for patients receiving a CYP3A4/PgP inducer). Similarly, in case of toxicity or trough

concentration above the target range, study treatment can also be decreased by one whole 2 mg tablet (or two 2 mg tablets for patients receiving a CYP3A4/PgP inducer).

Listings of all doses of the study treatment along with dose change reasons will be produced. Attention will be paid to patients in the everolimus 3 to 7 ng/ml treatment group, to identify separately the doses of everolimus tablets and placebo tablets that may have been taken.

3.7.1 Duration of study treatment exposure

The following algorithm will be used to calculate the duration of study treatment exposure for patients who took at least one dose of the study treatment:

Duration of exposure (days) = (date of last administration of study treatment) – (date of first administration of study treatment) + 1.

The duration includes any periods of temporary interruption of study treatment for any reason. "Date of last administration of study treatment" and "Date of first administration of study treatment" are defined in Section 2.1.4 and Section 2.1.5.

The duration of exposure will be summarized both categorically and continuously in units of weeks and/or months.

3.7.2 Cumulative dose

Cumulative dose is defined as the total dose given during the study treatment exposure, expressed in units of mg/m^2. For patients in the everolimus 3 to 7 ng/ml trough arm, total dose refers to the total amount of everolimus received during the trial (i.e., excluding placebo).

3.7.3 Dose intensity

Dose intensity (DI) for patients with non-zero duration of exposure is defined as follows:

DI (mg/m^2/day) = Cumulative dose (mg/m^2) / Duration of exposure (days).

3.7.4 Dose reductions, increases or interruptions

The number of patients with dose reductions, increases or interruptions, and the reasons, will be summarized by treatment group.

Note that since actual dose changes in the study are made in units of 2 mg, summaries across patients need to be interpreted with care, because for example, a 2 mg change is a higher relative change in dose for a patient with a BSA=0.6 than it is for a patient with BSA=2.0. Therefore, differences in dose changes between treatment arms may also be influenced by differences in BSA.

Interruption: An interruption is defined as one or more consecutive days on which no study medication is taken by the patient. A specific situation could arise for patients randomized to the everolimus arms, when the patient was still taking pills, but, due to an error in the kits used, all the pills were placebo. In that case the interruption of everolimus would not be reported with a reason in the eCRF since it was unknown. If that specific situation happens, then an interruption will be reported, along with the following reason: 'Only placebo tablets taken'. These cases will be identified programmatically, and the reason derived programmatically.

Reduction: A reduction is defined as a decrease in dose by 2 mg or more to a non-zero dose, even if this decrease has been directly preceded by an interruption. For example, in the sequence 8.0 mg/day – 0 mg/day – 6.0 mg/day, the 6.0 mg/day dose will be counted as a reduction.

Increase: An increase is defined as an increase in dose by 2 mg, even if this increase has been directly preceded by an interruption. For example, in the sequence 6.0 mg/day – 0 mg/day – 8.0 mg/day, the 8.0 mg/day dose will be counted as an increase.

3.8 Concomitant therapy

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) besides the study treatment (i.e., everolimus) that were administered to a patient preceding or coinciding with the study assessment period.

There will be one analysis for the double blind Core phase using the Safety Set, and a second analysis for all the data on everolimus from both the Core and Extension phases using the LTE Safety Set.

For the Core phase, concomitant includes therapy starting on or after the start date of double blind study treatment in the Core phase, or medications starting prior to the start date of study treatment in the Core phase and continuing after the start date. Therapy starting more than 30 days after the last day of study treatment in the Core phase will not be included, and neither will any therapy starting on or after the first day of everolimus given in the Extension phase. Prior concomitant therapy starts and ends before the start date of study treatment in the Core phase.

For the long-term evaluation of everolimus, concomitant includes therapy starting on or after the start date of everolimus, or medications starting prior to the start date of everolimus and continuing after the start date. Therapy starting more than 30 days after the last day of everolimus will not be included. Prior concomitant therapy starts and ends before the start date of everolimus: note that for patients randomized to the placebo arm and who subsequently received everolimus in the Extension phase, this refers to the entire Core phase and Baseline phase.

Concomitant medications entered into the database will be coded using the WHO (World Health Organization) Drug Reference List to allow for categorization by preferred term. In addition to categorizing medication data by preferred term, drugs are classified according to their ATC classification in order to present and compare how they are being utilized. The ATC classification allows for a summary of medications by a high-level common drug class.

Concomitant medications and significant non-drug therapies taken concurrently with study treatment will be listed and summarized by ATC class and preferred term by means of frequency counts and percentages. Concomitant AEDs received as background regimen will also be described separately and presented for both the Baseline and Core phases, including number of drugs received, type of drug received, and compliance to AED treatment.

Any prior concomitant medications (excluding prior anti-TSC therapy covered in Section 3.3.7 and prior antiepileptic therapy in Section 3.3.8) or significant non-drug therapy will be listed.

Concomitant medications with a specific impact on certain analyses will be identified prior to database lock, as defined in Section 2.4: background AED regimen, rescue medications and CYP3A4/PgP inducers and inhibitors. Separate summaries of these concomitant medications will be produced for the Core phase using the FAS population instead of the Safety Set. Medications will be listed and summarized by ATC class and preferred term by means of frequency counts and percentages.

3.9 Efficacy evaluation

The efficacy evaluation is based on the seizure diaries completed by patients on each day throughout the trial (see Section 2.5).

3.9.1 Primary efficacy endpoint

The primary efficacy endpoint is the reduction from baseline in the frequency of partial-onset seizures during the maintenance period of the Core phase. Two separate definitions of the primary efficacy endpoint will be used, owing to the differing regulatory preferences between Europe and USA for demonstrating efficacy of an antiepileptic medication. Whereas the EMA prefers **response rate** to be the primary variable, the FDA prefers **percentage reduction in seizure frequency**. It is understood that each Agency will use their preferred variable as the primary variable, with the other (non-primary) variable being used in a supportive analysis. As each Agency will only use their preferred primary variable to make a decision on the primary objective, the full alpha level can be used for each Agency's primary variable, without correction for multiplicity. This approach has been endorsed by both EMA and FDA.

3.9.1.1 Response Rate - definition

The following definitions are required to calculate response rate:

1. **Average weekly seizure frequency in the Baseline phase (SF_B)**
   Defined as:
   SF_B = 7 × number of partial-onset seizures recorded over the 8 week prospective Baseline phase ÷ number of non-missing seizure diary days in the 8 week prospective Baseline phase.

   Number of partial-onset seizures: The aggregate total of all seizures of codes IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 and IC6 (see Table 2-8), as reported in the "Investigator's Seizure Classification" eCRF.

   8 week prospective Baseline phase: The Baseline phase for the seizure diary is from the first day at which the diary is completed in screening through until the day before the date of randomization. Most patients are expected to be randomized on the day of the baseline visit, which is scheduled to be 8 weeks after the screening visit, in which case there will be a total of 56 days in the baseline phase seizure diary (assuming the diary is started on the day of the screening visit, as called for by the study protocol). However, due to late timing of the baseline visit for example, the baseline seizure diary could contain more than 56 days; or due to a delay in starting the diary at screening, the diary could contain less than 56 days. In either case the diary should be used in its entirety for this calculation of average weekly seizure frequency in the Baseline phase.

Number of non-missing seizure diary days: This is the number of days on which the number and types of seizures are known and documented in the "Investigator's Seizure Classification" eCRF. That is, it excludes days on which seizure information is unknown, coded as NA (not available) in the eCRF, which can arise when for example the patient or caregiver forget to count the seizures on a particular day.

2. **Average weekly seizure frequency in the maintenance period of the Core phase (SF_M)**

   Defined as:

   If patient does not discontinue during the 6 week titration period of the Core phase, then

   $SF_M = 7 \times$ number of partial-onset seizures recorded during the maintenance period of the Core phase ÷ number of non-missing seizure diary days in the maintenance period of the Core phase.

   Otherwise

   $SF_M = 7 \times$ number of partial-onset seizures recorded during the titration period of the Core phase ÷ number of non-missing seizure diary days in the titration period of the Core phase.

   That is, patients who discontinue prior to the maintenance period have seizure frequency determined using their data from the titration period, thereby assuring that all patients with seizure data in the Core phase have a value for SF_M.

   Number of partial-onset seizures: see point 1 above.

   Titration period of the Core phase: Defined as starting on the date of randomization until 6 weeks later, that is, from study day 1 until study day 42 inclusive; for patients who discontinue the study on or before study day 42, the titration period is considered to have ended on the last day of study treatment. Note that this rule will be the same for all patients, even if a visit is delayed and the final titration step occurs after study day 43.

   Maintenance period of the Core phase: Defined as starting at the beginning of week 7 of the Core phase (i.e., study day 43) and continuing until the last day of study treatment in the Core phase. For patients who complete the Core phase according to the scheduled visits, the last day of study treatment should be study day 126, corresponding to the end of week 18. Seizure counts beyond study day 126 should still be counted as long as double blind study treatment is continued in the Core phase.

   Number of non-missing seizure diary days: see point 1 above.

3. **Percentage reduction from baseline in average weekly seizure frequency during the maintenance period of the Core phase (%Red)**

   Defined as:

   $\%Red = 100 \times (SF_B - SF_M) \div SF_B$

   In the special case where SF_M is unknown, for example if a patient discontinues the trial without completing the patient seizure diary, then %Red is assigned equal to 0. (Note that SF_B is not expected to be unknown because the baseline seizure diary is required before a patient can be randomized.)

A responder is a patient with ≥ 50% reduction from baseline in average weekly partial-onset seizure frequency during the maintenance period of the Core phase, that is, when %Red ≥ 50. Response rate is the percentage of responders in a treatment group.

3.9.1.2 Percentage reduction in seizure frequency - definition

The FDA primary variable, percentage reduction in seizure frequency, is defined by the variable %Red defined in point 3 in Section 3.9.1.1 above.

3.9.1.3 Primary analysis – response rate

The primary analysis of response rate will compare each everolimus arm versus the placebo arm using Cochran-Mantel-Haenszel (CMH) chi-square tests (as described in Section 4.3). Each CMH test will be stratified by age subgroup at randomization, the randomization stratification factor, which consists of 4 subgroups: 1 to < 6, 6 to < 12, 12 to < 18, ≥18 years.

Multiplicity will be controlled via the Bonferroni-Holm procedure (Holm 1979), which is described below in Section 3.9.1.4 and in detail in Section 4.6. This method allows the statistical significance of each treatment comparison to be assessed whilst ensuring an overall family-wise alpha level of 2.5% one-sided.

Response rates will be provided with exact 95% confidence intervals (see Section 4.4), and the odds ratio (see Section 4.3) for each everolimus arm versus placebo will be obtained from logistic regression models stratified by age subgroup.

3.9.1.4 Hypothesis and test statistic - response rate

The statistical hypotheses are

$$H_{01}: RR_{EVE1} \leq RR_{PLB} \quad \text{versus} \quad H_{11}: RR_{EVE1} > RR_{PLB},$$

and

$$H_{02}: RR_{EVE2} \leq RR_{PLB} \quad \text{versus} \quad H_{12}: RR_{EVE2} > RR_{PLB},$$

where RR_{EVE1} is the probability of response on the everolimus 3 to 7 ng/ml arm, RR_{EVE2} is the probability of response on the everolimus 9 to 15 ng/ml arm, and RR_{PLB} is the probability of response on placebo.

The null hypothesis H_{01} will be tested against the alternative hypothesis H_{11} using the CMH test, with an associated one-sided p-value denoted by p_1. At the same time, the null hypothesis H_{02} will also be tested against the alternative hypothesis H_{12} using the CMH test, with an associated one-sided p-value denoted by p_2.

The Bonferroni-Holm procedure is used to ensure an overall family-wise Type I error of 2.5% one-sided, and it works as follows:

- If the smaller of the two one-sided p-values is greater than 0.0125, then neither of the two everolimus arms have a statistically significant benefit in the probability of response compared with placebo.
- Otherwise
  - if the smaller of the two p-values is less than or equal to 0.0125, then the everolimus arm associated with the smaller p-value has a statistically significant benefit in the probability of response compared with placebo.

  AND
  - If the larger of the two p-values is less than or equal to 0.025, then the everolimus arm associated with the larger p-value also has a statistically significant benefit in the probability of response compared with placebo.

3.9.1.5 Primary analysis – percentage reduction in seizure frequency

The primary analysis of percentage reduction in seizure frequency will compare each everolimus arm versus the placebo arm using rank analysis of covariance (rank ANCOVA, as described in Section 4.5). Each rank ANCOVA model will include baseline average weekly seizure frequency as covariate, and will be stratified by age subgroup at randomization, the randomization stratification factor, which consists of 4 subgroups: 1 to < 6, 6 to < 12, 12 to < 18, ≥18 years.

Multiplicity will be controlled via the Bonferroni-Holm procedure, which is described above in Section 3.9.1.4 and in detail in Section 4.6. This method allows the statistical significance of each treatment comparison to be assessed whilst ensuring an overall family-wise alpha level of 2.5% one-sided.

The median percentage reduction from baseline will be presented for each treatment group along with 95% bootstrap confidence intervals (see Section 4.7). Effect size will be assessed by calculating the difference between the median percentage reduction from baseline between

each of everolimus treatment arms and placebo using a stratified bootstrap methodology (stratified by the randomization stratification factor, see Section 4.7).

3.9.1.6 Hypothesis and test statistic - percentage reduction in seizure frequency

The statistical hypotheses are

H_{03}: $\mu_{EVE1} \leq \mu_{PLB}$ versus H_{13}: $\mu_{EVE1} > \mu_{PLB}$, and

H_{04}: $\mu_{EVE2} \leq \mu_{PLB}$ versus H_{14}: $\mu_{EVE2} > \mu_{PLB}$,

where μ_{EVE1} is the population mean of the reduction in seizure frequency for the everolimus 3 to 7 ng/ml arm, μ_{EVE2} is the population mean of the reduction in seizure frequency for the everolimus 9 to 15 ng/ml arm, and μ_{PLB} is the population mean of the reduction in seizure frequency for placebo.

The null hypothesis H_{03} will be tested against the alternative hypothesis H_{13} using rank ANCOVA, with an associated one-sided p-value for treatment effect denoted by p_3. At the same time, the null hypothesis H_{04} will also be tested against the alternative hypothesis H_{14} using rank ANCOVA, with an associated one-sided p-value for treatment effect denoted by p_4.

The Bonferroni-Holm procedure is used to determine whether p_3 and/or p_4 are statistically significant whilst maintaining an overall family-wise Type I error of 2.5% one-sided (see Section 3.9.1.4 for details).

3.9.1.7 Sensitivity and other supportive analyses

For the EMA primary variable of response rate, a supportive analysis will be the FDA primary analysis on percentage reduction in seizure frequency, and vice versa.

The primary analysis of both response rate and percentage reduction in seizure frequency will be repeated in the Per Protocol Set.

Descriptive statistics (n, mean, median, standard deviation, minimum and maximum) for average weekly seizure frequency in the Baseline phase (SF_B), in the maintenance period of the Core phase (SF_M), and percentage reduction from baseline (%Red), will be presented by treatment group in the FAS. The quantities for seizure frequency will also be summarized in units of 28 days instead of weekly, i.e., average number of partial-onset seizures per 28 days in the Baseline phase, and per 28 days in the maintenance period of the Core phase.

Sensitivity analyses will be conducted in the FAS to assess robustness of the primary analysis to the 4 following conditions presented in Table 3-2:

Table 3-2 Sensitivity analyses

Number	Description	Change compared to primary analysis Baseline phase	Change compared to primary analysis Core phase
1	Patient discontinuation before the end of Week 18 of the Core phase	None	Any patient discontinuing before day 126 will be classified as a non-responder, and for percentage reduction in

Number	Description	Change compared to primary analysis Baseline phase	Change compared to primary analysis Core phase
			seizure frequency, the reduction will be 0%.
2	Calculation of seizure frequency across the entire Core phase	None	Uses the average weekly seizure frequency over the entire Core phase (titration + maintenance)
3	'Best case' – Days with missing information considered as maximum seizure number for the Baseline phase and seizure-free for the Core phase	On days with missing information, the patient is considered to experience the same number of maximum daily seizures she/he has experienced during the baseline phase	Days with missing information considered to be seizure-free
4	'Worst case' - Days with missing information considered as maximum seizure number for the Core phase and seizure-free for the Baseline phase	Days with missing information considered to be seizure-free	On days with missing information, the patient is considered to experience the same number of maximum daily seizures she/he has experienced during the maintenance period

Below are the detailed calculation rules for the 4 sensitivity analyses:

1. Patient discontinuation before the end of Week 18 of the Core phase

   This sensitivity analysis will assume no change from baseline in seizure frequency for patients discontinuing before the end of Week 18 in the Core phase (i.e., discontinuing before day 126). This compares with the approach for the primary analysis where the observed seizure frequency calculated up to the day of discontinuation is used. Therefore, for response rate, any patient discontinuing before day 126 will be classified as a non-responder, and for percentage reduction in seizure frequency, the reduction will be 0%. Response rate and percentage change from baseline in seizure frequency will be recalculated for this sensitivity analysis and analyzed per the primary analysis.

2. Calculation of seizure frequency across the entire Core phase (i.e., titration period as well as maintenance period)

   This sensitivity analysis will use the average weekly seizure frequency over the entire Core phase, that is, from the day of randomization until the last day of study treatment in the Core phase. This compares with the primary analysis where titration period data will not be used to calculate average weekly seizure frequency, except if the patient discontinues on or before study day 42 (see Section 3.9.1.1 above).

   The variable to be used in the sensitivity analysis, labeled SF_{TM} is defined as follows:

   Average weekly seizure frequency in the entire Core phase (SF_{TM})

   Defined as:

   SF_{TM} = 7 × number of partial-onset seizures recorded over the entire Core phase ÷ number of non-missing seizure diary days in the entire Core phase.

   Number of partial-onset seizures: see point 1 in Section 3.9.1.1.

   Entire Core phase: From the day of randomization until the last day of study treatment in the Core phase. For patients who complete the Core phase according to the scheduled visits, the last day of study treatment should be study day 126, corresponding to the end of week 18. Seizure counts beyond study day 126 should still be counted as long as double blind study treatment is continued in the Core phase.

   Number of non-missing seizure diary days: see point 1 in Section 3.9.1.1.

   Response rate and percentage reduction in seizure frequency will be re-calculated using SF_{TM} instead of SF_M in the formula for %Red given in Section 3.9.1.1, and analyzed as per the primary analysis.

3. Calculation of seizure frequency assuming days missing data imputed to the maximum number of seizures experienced by the patient in the Baseline phase and to 0 seizures in the maintenance period of the Core phase.

   This sensitivity analysis will assume that the maximum number of seizures observed in the Baseline phase (SBmax) was observed on the days where data is reported as missing in the diary in this phase. For the Core phase, it will be assumed that 0 seizures were observed on the days where data is reported as missing in the diary. That is, there will be no missing seizure diary day in the Baseline and Core period phases. This compares with the primary analysis where days with missing seizure information are excluded from the calculations of the average weekly seizure frequency (see Section 3.9.1.1 above).

   The variables to be used in the sensitivity analysis, labeled SBmax, SF_{BMAX} and SF_{M0} are defined as follows:

   Daily maximum seizure count in the Baseline phase (SB_{MAX})

   Defined as:

   SB_{MAX} = maximum reported count of partial onset-seizures (see point 1 in Section 3.9.1.1) on one day during the Baseline phase (see point 1 in Section 3.9.1.1).

   Average weekly seizure frequency in the Baseline phase (SF_{BMAX})

Defined as:

SF_{B0} = 7 × number of partial-onset seizures recorded over the 8 week prospective Baseline phase ÷ number of days in the 8 week prospective Baseline phase.

Number of partial-onset seizures: The aggregate total of all seizures of codes IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 and IC6 (see Table 2-8), as reported in the "Investigator's Seizure Classification" eCRF. On days where information on seizure counts is missing in the diary, the patient will be considered to have experienced SBmax seizures.

8 week prospective Baseline phase: see point 1 in Section 3.9.1.1.

Number of days: This is the number of days of the Baseline phase. That is, it **includes** days on which seizure information is unknown.

Average weekly seizure frequency in the maintenance period of the Core phase (SF_{M0})

Defined as:

If patient does not discontinue during the 6 week titration period of the Core phase, then

SF_{M0} = 7 × number of partial-onset seizures recorded during the maintenance period of the Core phase ÷ number of days in the maintenance period of the Core phase.

Otherwise

SF_{M0} = 7 × number of partial-onset seizures recorded during the titration period of the Core phase ÷ number of days in the titration period of the Core phase.

That is, patients who discontinue prior to the maintenance period have seizure frequency determined using their data from the titration period, thereby assuring that all patients with seizure data in the Core phase have a value for SF_{M0}.

Number of partial-onset seizures: The aggregate total of all seizures of codes IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 and IC6 (see Table 2-8), as reported in the "Investigator's Seizure Classification" eCRF. On days where information on seizure counts is missing in the diary, the patient will be considered to have experienced 0 seizures.

Titration period of the Core phase: see point 2 in Section 3.9.1.1.

Maintenance period of the Core phase: see point 2 in Section 3.9.1.1.

Number of days: This is the number of days of the Maintenance (or Titration) phase. That is, it **includes** days on which seizure information is unknown.

Response rate and percentage reduction in seizure frequency will be re-calculated using SF_{B0} and SF_{M0} instead of SF_B and SF_M in the formula for %Red given in Section 3.9.1.1, and analyzed as per the primary analysis.

4. Calculation of seizure frequency assuming days with missing data imputed to the maximum number of seizures experienced by the patient in the maintenance period of the Core phase and to 0 seizures in the Baseline phase.

   This sensitivity analysis will assume that the maximum number of seizures observed in the Maintenance period of the Core phase (SCmax) was observed on the days where data is

reported as missing in the diary in this period. For the baseline phase, it will be assumed that 0 seizures were observed on the days where data is reported as missing in the diary. That is, there will be no missing seizure diary day in the Baseline and Core period phases. This compares with the primary analysis where days with missing seizure information are excluded from the calculations of the average weekly seizure frequency (see Section 3.9.1.1 above).

The variables to be used in the sensitivity analysis, labeled SCmax, SF_{B0} and SF_{MMAX} are defined as follows:

Daily maximum seizure count in the maintenance period of the Core phase (SC_{MAX})

Defined as:

If patient does not discontinue during the 6 week titration period of the Core phase, then

SC_{MAX} = maximum reported count of partial onset-seizures (see point 1 in Section 3.9.1.1) on one day during the maintenance period of the Core phase (see point 1 in Section 3.9.1.1).

Otherwise

SC_{MAX} = maximum reported count of partial onset-seizures (see point 1 in Section 3.9.1.1) on one day during the titration period of the Core phase (see point 2 in Section 3.9.1.1).

Average weekly seizure frequency in the Baseline phase (SF_{B0})

Defined as:

SF_{B0} = 7 × number of partial-onset seizures recorded over the 8 week prospective Baseline phase ÷ number of days in the 8 week prospective Baseline phase.

Number of partial-onset seizures: The aggregate total of all seizures of codes IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 and IC6 (see Table 2-8), as reported in the "Investigator's Seizure Classification" eCRF. On days where information on seizure counts is missing in the diary, the patient will be considered to have experienced 0 seizures.

8 week prospective Baseline phase: see point 1 in Section 3.9.1.1.

Number of days: This is the number of days of the Baseline phase. That is, it **includes** days on which seizure information is unknown.

Average weekly seizure frequency in the maintenance period of the Core phase (SF_{MMAX})

Defined as:

If patient does not discontinue during the 6 week titration period of the Core phase, then

SF_{MMAX} = 7 × number of partial-onset seizures recorded during the maintenance period of the Core phase ÷ number of days in the maintenance period of the Core phase.

Otherwise

SF_{MMAX} = 7 × number of partial-onset seizures recorded during the titration period of the Core phase ÷ number of days in the titration period of the Core phase.

That is, patients who discontinue prior to the maintenance period have seizure frequency determined using their data from the titration period, thereby assuring that all patients with seizure data in the Core phase have a value for SF_{MMAX}.

Number of partial-onset seizures: The aggregate total of all seizures of codes IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 and IC6 (see Table 2-8), as reported in the "Investigator's Seizure Classification" eCRF. On days where information on seizure counts is missing in the diary, the patient will be considered to have experienced SC_{MAX} seizures.

Titration period of the Core phase: see point 2 in Section 3.9.1.1.

Maintenance period of the Core phase: see point 2 in Section 3.9.1.1.

Number of days: This is the number of days of the Maintenance (or Titration) phase. That is, it **includes** days on which seizure information is unknown.

Response rate and percentage reduction in seizure frequency will be re-calculated using SF_{B0} and SF_{MMAX} instead of SF_B and SF_M in the formula for %Red given in Section 3.9.1.1, and analyzed as per the primary analysis.

3.9.1.8 Waterfall plots and cumulative plots

Waterfall graphs and cumulative plots will be used to investigate the effect of everolimus versus Placebo across a range of responder definitions. For the waterfall plot, the percentage reduction from baseline in seizure frequency during the maintenance period of the core phase will be presented for each patient. The cumulative plot will display the cumulative percentage of patients in each treatment arm for each individual value of the percentage reduction from baseline in seizure frequency observed in the maintenance period of the core phase.

3.9.2 Secondary efficacy endpoints

3.9.2.1 Seizure freedom

Seizure freedom is defined as not experiencing any partial-onset seizure (codes IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 and IC6 (see Table 2-8), as reported in the "Investigator's Seizure Classification" eCRF) during the maintenance period of the Core phase, where the maintenance period is from study day 43 until the last day of study medication in the Core phase. Note that this definition allows for patients to be classified as seizure-free despite experiencing partial-onset seizures during the titration period, and despite experiencing seizure types other than partial-onset during the maintenance phase. For patients who discontinue study treatment during the titration period, seizure freedom requires an absence of partial-onset seizures from study day 1 until the day of discontinuation. In terms of the variables defined in Section 3.9.1.1 above (without considering imputation rules for patients who discontinued the trial without completing the patient seizure diary), seizure freedom is when %Red = 100.

The seizure-free rates for each treatment arm in the FAS will be presented along with exact 95% confidence intervals, and the odds ratio for each everolimus arm versus placebo will be

derived from logistic regression models stratified by age subgroup. In addition, a sensitivity analysis will be performed where patients who discontinue at any time during the Core phase are assumed not to be seizure-free, even if no seizures had been reported before they discontinued.

3.9.2.2 Proportion of patients with at least 25% reduction in seizure frequency

This variable is equivalent to the primary endpoint of response rate except with a lower threshold of ≥ 25% reduction from baseline rather than ≥ 50%. In terms of the variables defined in Section 3.9.1.1 above (without considering imputation rules for patients who discontinued the trial without completing the patient seizure diary), a ≥ 25% reduction corresponds to %Red ≥ 25. The proportions of such patients in each treatment arm will be presented along with exact 95% confidence intervals. In addition, odds ratios for each everolimus arm versus placebo from logistic regression models stratified by age subgroup will be determined.

3.9.2.3 Distribution of reduction from baseline in seizure frequency

The distribution of reduction from baseline in seizure frequency will be categorized into the following six levels using the variable %Red defined in Section 3.9.1.1 above (without considering imputation rules for patients who discontinued the trial without completing the patient seizure diary):

%Red ≤ -25	exacerbation
-25 < %Red < 25	no change
25 ≤ %Red < 50	25% response
50 ≤ %Red < 75	50% response
75 ≤ %Red < 100	75% response
%Red = 100	seizure freedom

The proportions of patients in each category will be presented for each treatment arm.

3.9.2.4 Seizure-free days

Change from baseline in frequency of seizure-free days per 28 days will be presented for each treatment arm. The following definitions are required:

1. **Frequency of seizure-free days per 28 days during the Baseline phase ($SF0_B$)**
   Defined as:
   $SF0_B$ = 28 × number of seizure-free days recorded over the 8 week prospective Baseline phase ÷ number of non-missing seizure diary days in the 8 week prospective Baseline phase.

   Number of seizure-free days: The number of days on which the number and types of seizures were known and documented in the "Investigator's Seizure Classification" eCRF, and on which there were **no occurrences** of any of the seizure types that are part of the protocol definition of partial-onset (i.e., no occurrences of seizure codes IA1, IA2a, IB1, IB2, IB3, IC1, IC2, IC3, IC4, IC5 or IC6, as defined in Table 2-8).

8 week prospective Baseline phase: see point 1 in Section 3.9.1.1.

Number of non-missing seizure diary days: see point 1 in Section 3.9.1.1.

2. **Frequency of seizure-free days per 28 days in the maintenance period of the Core phase ($SF0_M$)**

   Defined as:

   If patient does not discontinue during the 6 week titration period of the Core phase, then

   $SF0_M$ = 28 × number of seizure-free days recorded during the maintenance period of the Core phase ÷ number of non-missing seizure diary days in the maintenance period of the Core phase.

   Otherwise

   $SF0_M$ = 28 × number of seizure-free days recorded during the titration period of the Core phase ÷ number of non-missing seizure diary days in the titration period of the Core phase.

   That is, patients who discontinue prior to the maintenance period have frequency of seizure-free days determined using their data from the titration period, thereby assuring that all patients with seizure data in the Core phase have a value for $SF0_M$.

   Number of seizure-free days: see point 1 above.

   Titration period of the Core phase: see point 2 in Section 3.9.1.1.

   Maintenance period of the Core phase: see point 2 in Section 3.9.1.1.

   Number of non-missing seizure diary days: see point 1 in Section 3.9.1.1.

Then the change from baseline in frequency of seizure-free days per 28 days is calculated as ($SF0_M$ - $SF0_B$), with values greater than 0 representing improvement over baseline.

Change from baseline in frequency of seizure-free days per 28 days will be summarized by treatment arm (mean, standard deviation, range). Mean differences between each everolimus arm and the placebo arm in change from baseline in frequency of seizure-free days per 28 days will be presented, along with 95% confidence intervals (See Section 4.9 for details on calculation of mean differences).

3.9.2.5 Treatment duration

Treatment duration is defined as the time from randomization until the date of permanent study treatment discontinuation (for any reason) at any time during the Core phase (i.e., titration period or maintenance period). Patients who complete the Core phase without discontinuing will have treatment duration censored on the last day of the Core phase. For the purposes of this analysis, a patient is considered to have completed the Core phase if still on double blind study treatment on study day 126 (where study day 1 is the randomization date).

The treatment duration distributions in each arm will be presented descriptively in the FAS using Kaplan-Meier curves (see Section 4.8.1), from which summary statistics will be determined, including the median treatment duration and the proportions of patients still on treatment at 6 and 12 weeks. These statistics will be given as point estimates with 95% confidence intervals. The hazard ratio (and two sided 95% confidence interval) for each everolimus arm versus placebo will be obtained from a Cox proportional hazards model stratified by age subgroup (see Section 4.8.2).

3.9.2.6 Long-Term Evaluation of Efficacy

A long-term evaluation of efficacy will be conducted using the LTE Efficacy Set, including all data on everolimus from both the Core phase and the Extension phase.

For patients randomized to one of the two everolimus arms, all data from the Core phase and the Extension phase will be used, whereas for patients randomized to placebo, only the Extension phase data will be included. In addition, the baseline seizure frequency for patients randomized to placebo will be obtained from the last 56 days prior to the start of everolimus, that is, from the last 8 weeks of the Core phase (see Table 3-2 below). There will be 4 different treatment groups displayed: everolimus 3-7 ng/ml, everolimus 9 to 15 ng/ml, placebo (i.e., extension phase data from patients previously randomized to placebo) and an overall everolimus arm.

The efficacy endpoints of response rate, percentage reduction from baseline in seizure frequency and change from baseline in seizure-free days will be determined at 12-weekly intervals throughout the duration of everolimus treatment (which for some patients could extend to more than two years). The analysis time points are shown in Table 3-2 below.

Table 3-3 Analysis time points for long-term evaluation of efficacy

Time point	Time window	Patients randomized to everolimus	Patients randomized to placebo
Baseline	The 8 weeks immediately prior to the start of everolimus	Study Weeks -8 to -1 (Baseline phase)	Study Weeks 11 to 18 (Last 8 weeks of Core phase)
Week 18	A 12-week window ending on the last day of the 18th week of everolimus	Study Weeks 7 to 18 (Maintenance period of Core phase)	Study Weeks 25 to 36 (Weeks 7-18 of Extension phase)
Week 30	A 12-week window ending on the last day of the 30th week of everolimus	Study Weeks 19 to 30 (Weeks 1 to 12 of Extension phase)	Study Weeks 37 to 48 (Weeks 19-30 of Extension phase)
Week 42	A 12-week window ending on the last day of the 42nd week of everolimus	Study Weeks 31 to 42 (Weeks 13 to 24 of Extension phase)	Study Weeks 49 to 60 (Weeks 31-42 of Extension phase)
Every 12 weeks thereafter			
Week =42+12×k, (where k=1,2,3,…)	A 12-week window ending on the last day of the (42+12×k)th week of everolimus	Study Weeks (31+12×k) to (42+12×k)	Study Weeks (49+12×k) to (60+12×k)
Note: Study Week 1 starts on the date of randomization			

The average weekly seizure frequency per time window (denoted SF_{TWi} for the i^{th} time window, i=1, 2, …), is defined as

SF_{TWi} = 7 × number of partial-onset seizures recorded during the i^{th} time window ÷ number of non-missing seizure diary days in the i^{th} time window,

and following the definitions provided in Section 3.9.1.1. The percentage reduction from baseline in seizure frequency at each time window, denoted $\%Red_{TWi}$ for the i^{th} time window (i=1, 2, …), is defined as

$\%Red_{TWi} = 100 \times (SF_{TW1} - SF_{TWi}) \div SF_{TW1}$

where SF_{TW1} is the average weekly seizure frequency in the baseline time window.

The frequency of seizure-free days per 28 days for the i^{th} time window, denoted $SF0_{TWi}$ (i=1, 2, …), is defined as

$SF0_{TWi}$ = 28 × number of seizure-free days recorded during the i^{th} time window ÷ number of non-missing seizure diary days in the i^{th} time window,

and the change from baseline in frequency of seizure-free days per 28 days is calculated as ($SF0_M$ - $SF0_B$), with values greater than 0 representing improvement over baseline.

Descriptive statistics (n, mean, median, standard deviation, minimum and maximum) for average weekly seizure frequency, percentage reduction from baseline and change from baseline in seizure-free days will be presented by treatment group at each time window, and median values with 95% bootstrap confidence intervals will be plotted over time. The proportions of patients with 50% response at each time window, defined where $\%Red_{TWi} \geq 50$, will also be shown for each treatment group, with 95% exact confidence intervals.

These analyses need to be interpreted with care because patients are permitted to change their background AED regimen during the Extension phase. So for example, a patient who has had an important reduction in partial-onset seizures during the Core phase, and who attributes that reduction to the use of everolimus, may reduce the dose or stop completely one or more of the background AEDs during the Extension phase. This in turn could lead to an increase in seizure frequency, which should not necessarily be interpreted as a sign of loss of efficacy of everolimus.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

3.10 Safety evaluation

The assessment of safety will be based mainly on the frequency of AEs and on the number of laboratory values that fall outside of pre-determined ranges. Other safety data including vital signs, neuropsychological scales and the Columbia Suicide Severity Rating Scale will also be considered.

Safety analyses will be performed on the Safety Set, and on the LTE Safety Set.

The safety summary tables will include only assessments collected no later than 30 days after study treatment discontinuation. All safety assessments will be listed and those collected later than 30 days after study treatment discontinuation will be flagged.

3.10.1 Adverse events data

3.10.1.1 Coding of AEs

AEs are coded using Medical Dictionary for Regulatory Activities (MedDRA) terminology.

3.10.1.2 Grading of AEs

The severity of AEs will be graded according to the Common Terminology Criteria for Adverse Events (CTCAE) version 4.03.

The CTCAE represents a comprehensive grading system for reporting toxicity in therapeutic oncology trials, but it has also been used outside of oncology (see http://wiki.nci.nih.gov) and was used in this study to be consistent with other studies of everolimus in TSC. CTCAE version 4.03 grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening and death. This grading system inherently places a value on the importance of an event although there is not necessarily proportionality among grades (Grade 2 is not necessarily twice as bad as Grade 1).

If CTCAE grading does not exist for an AE, grades 1 – 4 corresponding to the severity of mild, moderate, severe, and life-threatening will be used. CTCAE grade 5 (death) will not be used; death information will be collected on the "End of Treatment (Core Phase)", "End of Treatment (Extension Phase)" or "Study Completion Evaluation" eCRF pages.

3.10.1.3 General rules for AE Reporting

AE summaries will include all AEs starting or worsening on or after study day 1 (i.e., on or after the start date of study treatment), and starting no later than 30 days after the date of last

study treatment (see Section 2.1.8). Such events were labeled treatment emergent AEs in the study protocol, but will be referred to here as AEs in order to be consistent with the approach taken across other everolimus studies in TSC.

Pre-treatment AEs, defined as AEs starting or worsening between the time of signing informed consent and the day before starting study treatment, will only be summarized as part of the analysis of ongoing medical conditions – see Section 3.3.6. All AEs will be listed, and pre-treatment AEs will be flagged, as will post-treatment AEs, defined as AEs starting more than 30 days after the date of last study treatment.

AEs will be summarized by presenting the number and percentage of patients having at least one AE, and having at least one AE in each body system/primary system organ class, and for each preferred term using MedDRA coding. A subject with multiple occurrences of an AE will be counted only once in the AE category.

Separate AE summaries will be presented by primary system organ class, preferred term, and maximum CTCAE grade. A patient with multiple CTC grades for an AE will be summarized under the maximum (i.e., worse) CTC grade recorded for the event. In the summaries presented by grade, all AEs will be pooled regardless of whether they are CTC gradable or not. The frequency of CTC grade 3 and 4 AEs will be summarized separately.

Any information collected (e.g., CTC grades, relatedness to study drug, action taken, etc.) will be listed as appropriate.

3.10.1.4 AE summaries

The following AE summaries will be produced:

- Adverse events, regardless of study drug relationship, by primary system organ class and preferred term
- Adverse events, regardless of study drug relationship by preferred term and maximum CTC grade
- Adverse events with suspected relationship to study drug by primary system organ class, preferred term
- Adverse events, regardless of study drug relationship by primary system organ class, preferred term and maximum CTC grade
- On-treatment Deaths, by primary system organ class and preferred term
- All Deaths
- Serious adverse events, regardless of study drug relationship, by primary system organ class and preferred term
- Serious adverse events with suspected study drug relationship, by primary system organ class and preferred term
- Adverse events leading to study drug discontinuation, regardless of study drug relationship, by primary system organ class and preferred term
- Adverse events requiring dose adjustment or study-drug interruption, regardless of study drug relationship, by primary system organ class and preferred term

- Adverse events requiring additional therapy, regardless of study drug relationship, by primary system organ class and preferred term
- Non serious adverse events, regardless of study drug relationship, by primary system organ class and preferred term (>5%)
- Adverse events of special interest (AESI) (see Section 3.10.1.5) , regardless of study drug relationship, by grouping, preferred term and maximum CTC grade

- Adverse events of special interest (AESI), with suspected study drug relationship, by grouping, preferred term and maximum CTC grade
- For specific adverse events, analysis of time to event onset will be considered.

3.10.1.5 Grouping of adverse events of special interest (AESI)

Specific groupings of adverse events of special interest (AESI), labeled Safety Event Categories (SECs) in the study protocol, will be considered and the number of patients with at least one event in each grouping will be reported.

Specific groupings of AESI will be considered and the number of patients with at least one event in each grouping will be reported. Such groups consist of AEs for which there is a specific clinical interest in connection with everolimus treatment (i.e., where everolimus may influence a common mechanism of action responsible for triggering them) or AEs which are similar in nature (although not identical). The groups are defined at the project level and the latest version of the groupings based on the project-level information available at the time of the analysis will be used.

All AESI groupings are defined through a combination of Preferred Terms (PT), High Level Terms (HLT) or System Organ Classes (SOC). AE groupings definitions will be included in the appendix of the Clinical Study Report (CSR).

Amenorrhea

The proportions of patients having at least one event of amenorrhea will be summarized by treatment arm, using females from the Safety Set aged between 10 and 55while on treatment. Frequency of events, duration of the longest event, frequency of patients by duration of event and age at onset of first event will also be summarized.

3.10.1.6 Time to first onset of specific AEs

For selected groupings of AESIs or SOC, the following analysis of time to first occurrence will be considered for the double-blind Core phase.

The following AE groupings will be analyzed if at least 10% of patients have the corresponding event:

- Stomatitis (AESI)
- Infections and infestations (SOC)

All Treatment-Emergent Adverse Events (TEAEs) of interest (see Section 2.1.8) will be taken into account.

Time to first occurrence of an AE is defined as time from the start of study treatment to the date of first occurrence of an AE within the grouping considered in the analysis, i.e., time in days is calculated as (start date of first occurrence of AE) – (date of first dose of study treatment) +1.

A patient will be censored for time to first occurrence of an AE if:

- the patient discontinues the Core phase with no event
- the patient is still on-going at the end of the Core phase with no event

In the absence of an event, the censoring date applied will be the earliest from the following dates: end of double-blind study treatment + 30 days, start date of everolimus treatment in the Extension phase, data cutoff date, date of death.

Kaplan-Meier curves will be constructed by treatment group. Median time to first occurrence of AE together with the 95% confidence interval will be presented for each treatment group. The hazard ratio will be obtained from an unstratified Cox model.

The above analyses will be performed irrespective of AE grade, and may also be repeated for grade 3 or above.

3.10.2 Laboratory data

Laboratory assessments include hematology, biochemistry and urinalysis. Laboratory data from all sources (central and local laboratories) will be combined. The summaries will include all laboratory assessments collected no later than 30 days after the last treatment/exposure date (see Section 2.1.8). All laboratory assessments will be listed and those collected later than 30 days after the last treatment/exposure date will be flagged in the listings.

- All laboratory values will be converted into SI (Standard International) units and will be classified programmatically into CTC grades according to the NCI Common Terminology Criteria for Adverse Events (CTCAE) version 4.03.

- Conflict between normal range and grade definition: Because many institutions have differences for normal ranges of metabolic laboratory, and hematology values, the CTCAE often uses the terms 'Upper Limit of Normal (ULN)' and 'Lower Limit of Normal (LLN)' in lieu of actual numerical values. In some cases, an institution's LLN might be beyond the range specified for a Grade. In this case, the institutional limits of normal should take precedence over the CTCAE values: the laboratory value will still be taken as within normal limits and assigned a CTC grade of zero.
- For the few parameters having comparison to baseline in CTCAE grading definition (Fibrinogen, INR, Hemoglobin, Creatinine), the highest grade will be retained. In another words, in the particular case when a value is a grade x as per CTC grade definition based on threshold/ranges and also grade x+1 when comparing to baseline, Grade x+1 is retained.
- Grade 5 will not be used.
- For calcium, CTCAE grading is based on Corrected Calcium and not on Calcium. Corrected Calcium (CALC) can be calculated from Albumin and Calcium. Corrected Calcium (mg/dL) = Calcium (mg/dL) ‑ 0.8 [Albumin (g/dL)-4].

For laboratory tests where grades are not defined by CTCAE, results will be graded by the low/normal/high classifications based on laboratory normal ranges.

The following summaries will be produced for the laboratory data (by laboratory parameter):

- Number and percentage of patients by worst post-baseline CTC grade (regardless of the baseline status). Each patient will be counted only for the worst observed post-baseline grade. For laboratory parameters such as blood glucose where the patient can be graded for a decrease or an increase, the worst post-baseline CTC grade will be presented separately (i.e., hypoglycemia and hyperglycemia).
- Shift tables using CTC grades to compare baseline to the worst post-baseline value will be produced for hematology and biochemistry laboratory parameters that are CTC gradable.
- For laboratory parameters where CTC grades are not defined, shift tables to the worst post-baseline value will be produced using the low/normal/high classifications based on laboratory reference ranges.

3.10.2.1 Listings of laboratory values

The following listings will be produced for laboratory data:

- Listing of patients with laboratory values outside the laboratory reference ranges with values flagged to show the corresponding CTC grades and the classifications relative to the laboratory reference ranges
- A separate listing will display notable laboratory abnormalities (i.e., newly occurring CTCAE grade 3 or 4 laboratory toxicities).

3.10.2.2 Hepatitis B and C

All data collected on Hepatitis B and C will be listed.

The results obtained at baseline and at last examination will be summarized by treatment group. The number of patients followed for regular monitoring will also be provided.

3.10.3 Vital signs

Vital signs assessments are performed in order to characterize basic body function. Height (cm), weight (kg), body temperature (°C), pulse (beats per minute), systolic and diastolic blood pressure (mmHg) and respiration rate (breaths per minute) will be collected in eCRFs.

The criteria for clinically notable abnormalities are defined below in Table 3-4.

Table 3-4 Clinically notable values

Vital Sign		Patient age at randomization	
		< 18 years	≥ 18 years
Systolic BP	High	≥ 95th percentile of the age and height group[1]	≥ 180 mmHg and an increase ≥ 20 mmHg from baseline
	Low	≤ 5th percentile of the age and height group[1]	≤ 90 mmHg and an decrease ≥ 20 mmHg from baseline
Diastolic BP	High	≥ 95th percentile of the age and height group[1]	≥ 105 mmHg and an increase ≥ 15 mmHg from baseline
	Low	≤ 5th percentile of the age and height group[1]	≤ 50 mmHg and an decrease ≥ 15 mmHg from baseline
Body temperature	High	≥ 38.4 °C	≥ 39.1 °C
	Low	≤ 35.0 °C	
Weight	High	increase of ≥ 2 BMI-for-age percentile categories[2]	increase from baseline of ≥ 10%
	Low	decrease of ≥ 2 BMI-for-age percentile categories[2]	decrease from baseline of ≥ 10%
Pulse	High	≥ 120 bpm with increase from baseline of ≥ 15 bpm	
	Low	≤ 50 bpm with decrease from baseline of ≥ 15 bpm	

1: The 95th percentiles for blood pressure for the appropriate age and height groups are obtained from the NHLBI tables (http://www.nhlbi.nih.gov/guidelines/hypertension/child_tbl.htm).

2: Percentiles categories: P3, P5, P10, P25, P50, P75, P85, P90, P97 are obtained from the WHO Growth Charts (http://www.who.int/childgrowth/en/

The following summaries will be produced for each vital sign parameter:

- Number and percentage of patients with at least one post-baseline vital sign abnormality (in both directions, i.e., both elevated and below normal values).

In addition, the following two listings will be produced by treatment group:

- Patients with clinically notable vital sign abnormalities.
- All vital signs assessments will be listed by patient and vital sign parameter.

In both listings, the clinically notable values will be flagged and also the assessments collected later than 30 days after the date of last study treatment will be flagged.

3.10.4 Neuropsychological Scales

3.10.4.1 Vineland-II

The Vineland-II instrument (Vineland Adaptive Behavior Scales 2nd Edition) is for patients of any age and will be completed at baseline, end of Core phase, and then every 6 months (24 weeks) in the Extension phase plus at the end of treatment in the Extension phase. It will be used in only a subset of countries participating in the trial, where translated and validated

versions of the scale are available: Australia, Belgium, Canada, Colombia, France, Germany, Ireland, Italy, Japan, Netherlands, Poland, Spain, United Kingdom and United States.

There are two distinct versions of the Vineland-II being used in the study, namely the Survey Interview Form and the Parent/Caregiver Form. The Survey Interview Form is completed by the investigator during a semi-structured interview with the patient and/or caregiver, whereas the Parent/Caregiver Form is a questionnaire that is completed by the parent or caregiver directly. The two forms differ only in the method of administration, although the wording is different on the Parent/Caregiver Form for ease of understanding. The same type of form that is used at the baseline assessment should be used thereafter at all subsequent assessments.

Each form contains 433 items that are scored 0 (never), 1 (sometimes) or 2 (usually). The items are organized into the five domains of Communication, Daily Living Skills, Socialization, Motor Skills and Maladaptive Behavior. The five domains consist of 15 subdomains as shown in Table 3-5.

Table 3-5 Content of Vineland-II

Domain (number of items)	Subdomains (number of items)
Communication (99)	Receptive (20), Expressive (54), Written (25)
Daily Living Skills (109)	Personal (41), Domestic (24), Community (44)
Socialization (99)	Interpersonal Relationships (38), Play and Leisure Time (31), Coping Skills (30)
Motor Skills (76)	Gross (40), Fine (36)
Maladaptive Behavior (50)	Maladaptive Behavior Index – Internalizing (11), Maladaptive Behavior Index – Externalizing (10), Maladaptive Behavior Index – Other (15), Maladaptive Behavior Critical Items (14)

Within each domain, the subdomains yield so-called *v*-scale scores that sum to yield the domain composite scores. An overall Adaptive Behavior Composite (ABC) score is obtained by combining the first four domain scores for patients aged less than 7 years, or the first 3 domain scores for patients aged 7 or older.

The raw scores, *v*-scale scores for subdomains, standard scores for domains and overall Adaptive Behavior score will be obtained using the Vineland-II Survey Forms ASSIST™ software, which provides scores based on the Vineland-II Survey Forms Manual (Sparrow et al. 2005). Individual item answers, as collected during administration of the questionnaire, will be entered in the software by an external vendor. Subsequently, scores obtained from the software will then be printed and results entered in a dataset. This dataset containing the derived scores (v-scale scores for subdomains, domain standard scores and ABC score) will then be transferred to the Novartis systems for analysis.

The 14 Maladaptive Behavior Critical Items do not yield a score and will only be listed. The 15 Maladaptive Behavior Index – Other items do not yield a v-scale score and will not be analyzed separately. They will only be used for the derivation of the Maladaptive Behavior Index Score, and will also be listed.

Compliance to the schedule of administration of the Vineland-II will be summarized for each treatment arm in the FAS at baseline and end of Core phase using the following categories:

- yes, fully completed

- yes, partly completed
- no

The precise rules to determine completeness of the forms will be described in RAP M8.

For the long-term evaluation of everolimus, the same categories will be used to summarize compliance in the LTE Safety Set at baseline, end of Core phase and then every 6 months (24 weeks) in the Extension phase plus at the end of treatment in the Extension phase.

Descriptive statistics (n, mean, standard deviation, minimum and maximum) for the overall Adaptive Behavior Composite and domain/subdomain *v*-scale scores will be presented at baseline and at end of Core phase by treatment group in the Safety Set, using the time windows defined in Section 2.1.9.1. The data will also be presented separately based on the type of form that is used (Survey Interview Form and Parent/Caregiver Form).

Change from baseline at the end of the Core phase in these scores will also be summarized. For this analysis, only patients who used the same form at baseline and end of Core phase will be included.

A similar descriptive approach will be used for the long-term evaluation of everolimus, including change from baseline in each time window through to the end of the Extension phase.

During the conduct of the study, it became apparent through blinded data review that questionnaires (caregiver forms in particular) were not being filled out correctly in some cases, preventing the determination of raw scores for some subdomains. For each subdomain, a basal and a ceiling item must be established in order to determine the patient's score. The basal item is defined when a series of 4 consecutive 2's (usually) is determined in the subdomain and the ceiling item is defined when a series of 4 consecutive 0's (never) is found. The Parent/Caregiver Form identifies a starting point answering subdomain items, based on patient's chronological age. In patients with cognitive disability, chronological age may not correlate with developmental age and an earlier starting point would be more appropriate. In that case, the basal item would be established at lower levels than what is expected for this age range. This would require answering items targeting skills associated with younger patients in an effort to establish a basal item. The same issue could be seen for the ceiling item, less commonly, where it might be needed to answer items from age categories above the patient's chronological age. It is possible, therefore, that because questionnaires were completed based on a chronological age-starting point, which is often greater than the developmental age in the largely disabled population enrolled in this study, basal and ceiling items were often not established properly.

As a consequence, subdomain raw scores were often missing and, subsequently, subdomain *v* -scale scores, domain standard scores and ABC scores were missing as well. To mitigate the impact of many missing data points, an attempt was made to retrospectively collect the missing information and increase the likelihood of establishing basal and ceiling items in patients.

It is recognized that self-reported (or third-party reported) outcomes, can be subject to recall bias, and therefore, the primary analyses described above will be conducted including only data collected on 'real-time', as initially planned in the protocol. Some analyses will be

repeated on a dataset that includes retrospectively collected data and identified as 'real-time and retrospective' in the outputs. These supportive analyses will include data from retrospective collection, only when the retrospective collection yielded a score that was previously missing.

Retrospectively collected data (individual items answers and scores) will be flagged in the database via a different CRF page name, to allow the conduct if the primary and supportive analyses as described above.

3.10.4.2 Wechsler Nonverbal

The Wechsler Nonverbal instrument (WNV: Wechsler Nonverbal Scale of Ability) is for patients aged 4 to 21 years at baseline (i.e., ≥ 4 and < 22), and will be completed at baseline, end of Core phase, and then every 6 months in the Extension phase plus at the end of treatment in the Extension phase. It is a non-verbal measure of general ability, with pictorial instructions and no requirement for respondents to speak, and will be used in all sites across the study.

The full version of the WNV consists of four tests for patients less than 8 years old, namely Matrices, Coding, Object Assembly and Recognition, and four tests for patients aged 8 to 21, which are Matrices, Coding, Spatial Span and Picture Arrangement. The brief version of the WNV includes only Matrices and Recognition for patients under 8, and Matrices and Spatial Span for patients aged 8 to 21. Taking into account the time required for this and other scales that will be collected at the same time (i.e., Vineland-II, quality of life), it was decided to use the brief version of WNV for this trial, but to also perform the Coding test.

Each test is scored by the investigator, and the raw scores are recorded in the eCRF, i.e., raw scores for Matrices and Recognition for patients aged under 8 years, and for Matrices and Spatial Span for patients aged 8 to 21. The raw score for Coding is also obtained, along with the completion time in seconds, for all patients aged 4 to 21. In addition, the following four quantities can be determined from the Spatial Span test for patients aged 8 to 21: Spatial Span Forward (SSpF), Spatial Span Backward (SSpB), Longest Spatial Span Forward (LSSpF) and Longest Spatial Span Backward (LSSpB).

Based on the raw scores from the subtests, standardized z-scores will be determined for all of the subtests/quantities defined above.

For each of the subtests, the mean (b) and standard deviation (Sb) of the raw scores will be computed using data from all subjects at baseline. Then, the z-score for a subtest for each subject at each visit (including baseline visit) will be computed from the raw score (X) as follows:

$$Z = (X - b)/Sb .$$

Z follows a standard normal distribution.

A composite WNV score (W) for each subject at each visit will be computed by summing up the Z-scores of the 3 subtests of the WNV collected in this study (i.e., Matrices, Recognition and Coding for patients aged under 8 years; Matrices, Spatial Span and Coding for patients aged 8 to 21).

Compliance to the schedule of administration of the WNV will be summarized for each treatment arm in the FAS at baseline and end of Core phase using the following categories, which will be determined programmatically based on presence/absence of raw scores in the eCRF:

- yes, fully completed
- yes, partly completed
- no

For the long-term evaluation of everolimus, the same categories will be used to summarize compliance in the LTE Safety Set at baseline, end of Core phase and then every 6 months in the Extension phase plus at the end of treatment in the Extension phase.

Descriptive statistics (n, mean, standard deviation, minimum and maximum) for the composite WNV score, and the individual test *z*-scores, will be presented at baseline and at end of Core phase by treatment group in the FAS, using the time windows defined in Section 2.1.9.1. Change from baseline at the end of the Core phase in these scores will also be summarized.

A similar descriptive approach will be used for the long-term evaluation of everolimus, including change from baseline in each time window through to end of the Extension phase.

3.10.5 Columbia Suicide Severity Rating Scale

The electronic Columbia Suicide Severity Rating Scale (eC-SSRS) will be completed by patients at all visits except at the PK visits (weeks 1, 3, 5 and 19) and the End of Study visit. It will be administered by an Interactive Voice Response System (IVRS), with the patient making the telephone call while at the study site. For patients under 12 years of age, the suicide scale will be answered on their behalf by the parent or caregiver.

The scale consists of five categories of suicide ideation (Categories 1 to 5) and five categories of suicide behavior (Categories 6 to 10) each answered either yes or no:

Suicide Ideation:

Category 1 – Wish to be Dead
Category 2 – Non-specific Active Suicidal Thoughts
Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Category 5 – Active Suicidal Ideation with Specific Plan and Intent

Suicide Behavior:

Category 6 – Preparatory Acts or Behavior
Category 7 – Aborted Attempt
Category 8 – Interrupted Attempt
Category 9 – Actual Attempt (non-fatal)
Category 10 – Completed Suicide

Self-injurious behavior without suicidal intent is also an eC-SSRS outcome (although not suicide-related) and will also be answered either yes or no.

Three composite endpoints based on the above categories will be used:

- Suicidal ideation is defined as a "yes" at any time on or after the first day of study treatment to any one of the five suicidal ideation categories (Categories 1 to 5).
- Suicidal behavior is defined as a "yes" at any time on or after the first day of study treatment to any one of the five suicidal behavior categories (Categories 6 to 10).
- Suicidal ideation or behavior is a "yes" answer at any time on or after the first day of study treatment to any one of the ten categories.

In addition, the eC-SSRS data will be mapped to the Columbia Classification Algorithm for Suicide Assessment (C-CASA) as per FDA guidance on suicidality, as follows:

1. Completed Suicide: eC-SSRS Category 10.
2. Suicide Attempt: eC-SSRS Category 9.
3. Preparatory Actions Toward Imminent Suicidal Behavior: eC-SSRS Categories 6 to 8.
4. Suicidal Ideation: eC-SSRS Categories 1 to 5.
5. Self-Injurious Behavior without Suicidal Intent: taken directly from eC-SSRS.

Compliance to the schedule of administration of the eC-SSRS will be summarized for each treatment arm in the Safety Set at each of the time windows defined in Section 2.1.9.4. Compliance will be determined programmatically based on the responses in the eC-SSRS, and will use the following categories:

- yes, fully completed
- yes, partly completed
- no

A call is considered as fully completed if answers to all of the questions related to the categories 1 to 9 above have been given at the IRT by the respondent. Please note that categories 3, 4 and 5 are only inquired in case of a positive answer to question 2; therefore, calls with a negative answer to question 2 and missing results to questions 3, 4 and 5 in the database will be considered as fully completed calls.

In the analysis, some calls to the IRT will be considered as invalid and results not included in the summary tables. These calls will be flagged with a '*' as invalid calls in the listings. These calls are defined "invalid" when the patient or caregiver calling the IRT system recording the answers to the questionnaire hangs up before completing the call. If the patient or caregiver calls back the same day and re-answer all questions and completes the call, then the first incomplete call is flagged as invalid and is not taken into account in the analyses. This situation typically arises when the patient/caregiver answered incorrectly one question and therefore hangs-up and calls back to correct the answers. Both calls will be listed. Please note that an incomplete call not followed by a complete call on the same day is still considered valid and is included in the analyses.

For the long-term evaluation of everolimus, the same categories will be used to summarize compliance in the LTE Safety Set at each of the time windows. Note that for placebo patients who start everolimus in the Extension phase, their compliance data will be shown using the Extension phase time windows (i.e., from Week 22 onwards).

A data listing will be prepared containing all patients in the Safety Set who answered "yes" during the Baseline or Core phases to any of the Categories 1 through 10, or who answered "yes" to the outcome "Self-injurious behavior without suicidal intent". The listing will contain all visits for such patients from screening through to the end of the Core phase, and the responses to all ten Categories and the additional outcome on self-injurious behavior will be shown, along with the C-CASA score. A similar listing will be presented using the LTE Safety Set, including patients who answered yes during Baseline, Core or Extension phases.

The proportions of patients in each treatment group in the Safety Set with suicidal ideation, suicidal behavior and suicidal ideation or behavior, at any time during the Core phase, will be presented, along with the worst categories of suicidal ideation and suicidal behavior (where worst means the highest numbered category). The same analysis will be performed using the C-CASA scores of completed suicide, suicide attempt, preparatory actions toward imminent suicidal behavior, suicidal ideation and self-injurious behavior without suicidal intent. These analyses will be repeated using the LTE Safety Set across the Core and Extension phases.

Upon release of protocol amendment 2, investigators were required to discuss episodes of self-injury and changes in the patient's mood and/or behavior with the patient and caregiver, for all patients. For patients ≥ 13 years of age who are not able to complete the eC-SSRS scale via IRT, the investigator was to proactively assess the patient for the presence of suicidality. Any positive response had to be further investigated, and appropriate medical treatment initiated and recorded as an adverse event on the AE eCRF.

The summary of this assessment was recorded in the eCRF and data will be listed.

3.10.6 Other safety data

Data from other tests (e.g., electrocardiogram) will be listed, notable values will be flagged, and any other information collected will be listed as appropriate.

All assessments collected later than 30 days after the last date of study treatment (see Section 2.1.8) will be flagged in the listings.

3.11 Other test data

3.11.1 Growth Data

Growth data will only be analyzed as part of the long-term evaluation of everolimus. This is because the 18 week Core phase is not considered a long enough period of time to observe changes in growth. Also, analyses will be restricted to patients under the age of 18, in whom height/weight growth is still likely to be taking place. Therefore, all analyses will use patients from the LTE Safety Set who were under 18 years of age on the start date of everolimus.

Height and BMI will be summarized at 24-week intervals before and after starting everolimus, using the standard deviation score (SDS, also called z-score), velocity and velocity SDS. The relevant height and BMI values for each 24-week period are defined using time windows, as defined in Section 2.1.9.5. BMI will be calculated as weight (in kg) / squared height (in m). The z-scores will allow identification of potential outliers.

SDS will be calculated using the current formulae provided by the WHO as follows:

1. Calculate $z_{ind} = \frac{\left(\frac{X}{M}\right)^{L} - 1}{LS}$
2. If $|z_{ind}| \leq 3$, $SDS = z_{ind}$
   If $z_{ind} > 3$, $SDS = 3 + (X - SD3pos) / SD23pos$
   If $z_{ind} < -3$, $SDS = -3 + (X - SD3neg) / SD23neg$

where:

- *X* is height in centimeters or BMI in kilograms/m^2,
- *L*, *M* and *S* are height or BMI-, sex- and age-specific reference values from the WHO Growth Charts.
- SD3*pos* is the cutoff 3SD calculated by the LMS method:
  $SD3pos = M * (1 + LS*3)^{1/L}$
- SD3*neg* is the cutoff -3SD calculated by the LMS method:
  $SD3neg = M * (1 + LS*(-3))^{1/L}$
- SD23*pos* if the difference between the cutoffs 3SD and 2SD:
  $SD23pos = M * (1 + LS*3)^{1/L} - M * (1 + LS*2)^{1/L}$
- SD23*neg* if the difference between the cutoffs -2SD and -3SD:
  $SD23neg = M * (1 + LS*(-2))^{1/L} - M * (1 + LS*(-3))^{1/L}$

Height-for-age and BMI-for-age L, M and S reference values for males and females are available under http://www.who.int/childgrowth/standards/en/ (for patients aged between 0 to 5 years old) and http://www.who.int/growthref/en/ (for patients aged between 5 to 19 years old). These correspond to the latest available international references available at this time and described in the 2007 Bulletin of the World Health Organization (Mercedes de Onis et al 2007). The age category immediately above the patient's exact age should be used. SDS is actually a Z score that measures the distance from the population mean in units of standard deviations. That is, SDS < 0 refers to values lower than the population mean, and for example SDS ≤ -1.645 refers to values in the lowest 5%. (The usual percentile more commonly used in the clinical practice can be derived from the Z-score by a normal distribution).

Height velocity is defined as follows:

Height velocity (cm/6-months) = (height in time window *k* – height in time window *k-1*)
÷ ([assessment date in time window *k* – assessment date in time window *k-1*] ÷ [365.25/2]),

and similarly for weight velocity.

Velocity SDS is calculated as (velocity – mean) / SD, where mean and SD are obtained as the height-, weight-, sex- and age-specific values in Tables 3 to 8 in Baumgartner (1986), where the age category immediately above the patient's exact age (at the assessment date in time window *k*) should be used. Velocity SDS will only be calculated for time window *k* if data also exists for time window *k-1*, since calculating across multiple units of 6 months requires more than one reference value to be taken into account.

Height/weight SDS and velocity SDS will be summarized using descriptive statistics (mean, standard deviation, range) for each time window (before and after starting everolimus allowing informal comparison of growth data), as well as by presenting number of patients

with SDS values lower/higher than $5^{th}/95^{th}$ percentiles respectively. Box plots will also be plotted for each time window. All height/weight SDS, velocity and velocity SDS data will be listed, and values of SDS and velocity SDS outside of the central 95% of population values will be flagged as either High (SDS $\geq$ 1.645) or Low (SDS $\leq$ -1.645).

Comparison of height at 18 years with mid-parental heights will be performed, by presenting the number of patients with a height higher/lower than expected. High/Low values are defined as values more extreme than mid-parental height +/- 8.5 cm. Mid-parental height= ((father's height – 13) + mother's height) ÷ 2 for girls, and ((mother's height + 13) + father's height) ÷ 2 for boys (Tanner et al. 1970).

3.11.2 Puberty Stage

Puberty stage will only be analyzed as part of the long-term evaluation of everolimus on the LTE Safety Set. .

Tanner Stage includes two components for boys, namely testis and pubic hair, and two components for girls: breast development and pubic hair. It is expected that data will become available during the trial on a proportion of patients as they go through puberty attaining higher levels of the Tanner Stage. For the age at which Tanner Stages 2-5 are achieved, age at thelarche (females), age at menarche (females) and age at adrenarche (males), summary statistics from Kaplan-Meier distributions will be determined, including the median age and the proportions of patients reaching these milestones at some given ages (see Section 4.8.1). These statistics will be given as point estimates with 95% confidence intervals. The percentage of patients who will reach Tanner Stage 2 during the study will be calculated among the number of patients with Tanner Stage 1 at Baseline (prior to the start of everolimus). The percentage of patients who will reach Tanner Stage 3 during the study will be calculated among the number of patients with Tanner Stage 1 or 2 at Baseline.

Similar rules will be applied for the age at Tanner Stages 4 and 5. Age at thelarche, age at menarche and age at adrenarche will be assessed among patients who have not yet reached these development milestones at baseline.

Potential delayed puberty in girls is defined as failure to attain Tanner Stage 2 (for both breast development and pubic hair) by age 13, or absence of menarche by age 15 or within 5 years of attainment of Tanner Stage 2 (Fenichel et al. 2012). Potential delayed puberty in boys is defined as failure to attain Tanner Stage 2 (for both testis and pubic hair) by age 14 (Crowley et al. 2012). Rates of potential delayed puberty will be presented for boys and girls separately, along with 95% confidence intervals, on the population at risk of delayed puberty at baseline. Note that the denominator for analyses of delayed puberty excludes patients who were identified as having delayed puberty at baseline (i.e., prior to starting everolimus). Potential cases identified through this algorithm, will be then clinically reviewed by assessing all available information in order to conclude the clinical relevance of the delay.

Blood samples for endocrine testing will be obtained at week 14, end of treatment in the Core phase, every 12 weeks through the Extension phase and at end of treatment in the Extension phase. Parameters measured will include testosterone, follicle stimulating hormone, luteinizing hormone and estradiol (girls only). The data will be listed.

3.12 Patient reported outcomes

3.12.1 Quality of life

Quality of life will be assessed using three baseline-age-specific questionnaires: the Quality of Life in Childhood Epilepsy (QOLCE) for patients aged ≤ 10 years, the Quality of Life in Epilepsy Inventory for Adolescents-48 (QOLIE-AD-48) for patients aged 11 to ≤ 17 years, and the Quality of Life in Epilepsy Inventory-31-Problems (QOLIE-31-P) for patients aged ≥ 18 years.

Each questionnaire will be completed at baseline, at end of Core phase (or end of treatment if patient discontinues), and then at the end of treatment in the Extension phase. The same questionnaire that is used at baseline should be used at the post-baseline visits, even if the patient's age exceeds the upper age limit, so that changes from baseline can be calculated.

Details of each of the three questionnaires follow.

QOLCE

The QOLCE is completed by the patient's parent or caregiver. It consists of 16 subscales (13 multi-item scales and 3 single item scales) and one Overall Quality of Life Score. The subscale score for each individual is calculated as the mean of the items belonging to the subscale. The Overall Quality of Life Score is computed by adding each subscale score for each individual and then dividing by 16. Higher scores correspond to improved quality of life.

The 16 subscales are as follows, where the numbers and letters following each subscale name refer to the section and item numbers in the questionnaire:

Quality of life item:	2.1.
Physical restrictions:	3.1. a, b, c, d, e, f, g, h, i, j.
Energy/fatigue:	3.2. a, b
Depression:	4.1. a, d, e, l
Anxiety:	4.1. g, j, n, o, p
Control/helplessness:	4.1. c, f, h, i
Self-esteem:	4.1. k, m, q, r, s
Attention/concentration:	5.1. a, d, e, f, g
Memory:	5.1. j, k, l, m, n, o
Language:	5.1. p, q, r, s, t, u, v, w
Other Cognitive:	5.1. b, c, h
Social interactions:	6.1. b, c, d, f, h
Social activities:	6.1. a, e, and 6.2.
Stigma item:	6.1. i
Behaviour:	7.1. a, b, c, f, g, h, i, j, k, l, m, o, p, q, r, t
General health item:	8.1.

The coding of each item and subscale will be given in detail in Module 8 of the RAP documentation, based on the publication by Sabaz et al. 2003. Higher scores correspond to improved quality of life. An Overall Quality of life score can be obtained by averaging the 16

subscale scores. There has been doubt expressed over the reliability and validity of the tool in patients under 6 years of age for the cognitive subscales: Attention/concentration, Memory, Language, Other Cognitive (Sabaz et al. 2003). For these patients, the Overall Quality of life score will be obtained as the average of the remaining 12 subscales.

QOLIE-AD-48

The QOLIE-AD-48 must be completed by the patient, and not by the parent or caregiver. It contains 48 items in 8 subscales as follows, where the numbers following each subscale name refer to the item numbers in the questionnaire:

Epilepsy impact (12 items):	7, 26, 27, 29, 30, 31, 32, 33, 34, 35, 37, 48.
Memory/concentration (10 items):	9, 13, 14, 15, 16, 17, 18, 19, 20, 21.
Attitudes toward epilepsy (4 items):	44, 45, 46, 47.
Physical functioning *(5* items):	3, 4, 5, 6, 8.
Stigma (6 items):	38, 39, 40, 41, 42, 43.
Social support (4 items):	22, 23, 24, 25.
School behavior (4 items):	10, 11, 12, 28.
Health perceptions (3 items):	1, 2, 36.

The coding of each item and subscale will be given in detail in Module 8 of the RAP documentation, based on the publication by Cramer et al. 1999. Higher scores correspond to improved quality of life. An overall quality of life score is obtained by summing a linear combination of the subscale scores, where each subscale is multiplied by a relative weight that is provided in the publication.

QOLIE-31-P

The QOLIE-31-P must be completed by the patient, and not by the parent or caregiver. It contains 39 items, of which a total of 30 are used to make up 7 different subscales as follows, where the numbers following each subscale name refer to the item numbers in the questionnaire:

Seizure Worry (5 items):	30, 31, 32, 33, 34.
Overall Quality of Life (2 items):	1, 36.
Emotional well-being (5 items):	7, 8, 9, 10, 11.
Energy/fatigue (4 items):	2, 3, 4, 5.
Cognitive (6 items):	19, 20, 21, 22, 23, 24.
Medication effects (3 items):	26, 27, 28.
Social function (5 items):	13, 14, 15, 16, 17.

A further 7 items (numbers 6, 12, 18, 25, 29, 35, 37 in the questionnaire) ask about distress, one for each of the 7 subscales. These distress scores provide the weights in a linear combination of the 7 subscale scores to give an overall quality of life score. Finally, item 38 assesses the overall health of the patient, and item 39 requires patients to rank in order of importance the 7 subscale domains in the list above. The coding of each item will be given in detail in Module 8 of the RAP documentation, based on the publication by Cramer et al. 2003 and the QOLIE-31-P Scoring Manual (QOLIE Development Group 2013).

QOL statistical analysis

There will be a common approach to analyzing each of the three questionnaires. The exception being that for the QOLCE cognitive subscales (Attention/concentration, Memory, Language, Other Cognitive), only patients aged 6 years and over at the baseline assessment will be included. First, compliance to the schedule of administration of each questionnaire will be summarized for each treatment arm in the FAS at baseline and end of Core phase using the following categories:

- yes, fully completed
- yes, partly completed
- no

For the long-term evaluation of everolimus, the same categories will be used to summarize compliance in the LTE Efficacy Set at baseline, end of Core phase and end of Extension phase. In addition, for all three QOL questionnaires, the number of items that can be missing for a given patient for a subscale score to still be calculated will be determined based on the 50% or "half-scale" rule. That is, for all three QOL questionnaires, subscores are calculated for a given patient only if at least 50% of the items in the subscale are completed; otherwise, the score for that subscale is set to missing for that patient. Thus, for example, the Epilepsy Impact domain of the QOLIE-AD-48 consists of 12 items, and so a specific patient must answer at least 6 of those items for their Epilepsy Impact score to be calculated. If a patients answers less than 6 items then the subscale score is set to missing. For the QOLIE-AD-48 and the QOLIE-31-P, the total score should not be calculated if more than one scale score is missing. If one scale is missing, the total score is calculated based on the subscale scores that are completed. In addition, the following specific rule must be applied for the derivation of the QOLIE-AD-48 total score: the "Epilepsy impact" subscale score (with a weight of 0.31 in the total score) is always required. For the QOLCE, if less than 50% of the subscale scores are present then the total score is also set to missing.

Descriptive statistics (n, mean, standard deviation, minimum, maximum) for the overall quality of life score and subscale scores will be presented at baseline and at end of Core phase by treatment group in the FAS. Change from baseline at the end of Core phase in these scores will also be summarized.

Change from baseline to end of Core phase in the overall quality of life scores will be analyzed using an analysis of covariance (ANCOVA) model including terms for treatment and baseline overall quality of life score. The differences in least square means between each everolimus arm and the placebo, and the corresponding two-sided 95% confidence interval will be presented.

A similar descriptive approach will be used for the long-term evaluation of everolimus, including change from baseline at end of Core phase and End of Extension phase.

3.13 Pharmacokinetic analysis

This study includes PK analyses for everolimus concentration, and also for concentrations of 12 different AEDs.

3.13.1 Everolimus PK

3.13.1.1 General principle

In this study, measurement of everolimus levels at trough (Cmin) is scheduled for weeks 1, 3, 5, 10, 14 and 18 during the Core phase, and at weeks 19, 22, 26, and 30 , and every 12 weeks thereafter during the Extension phase. Additional everolimus PK samples will be collected 2 weeks after any change in the dose of study medication, or 2 weeks after any change in use of concomitant CYP3A4/PgP inhibitors or inducers.

Blood samples for the pre-dose sample (Cmin) should be obtained prior to dose administration on the same treatment day and at 20-28 hours after the previous dose. Cmin should be collected at steady state, which means that no dose interruption or dose changes should have occurred in the previous 4 days. Further, the Cmin value will not be reliable if the patient had vomited within 4 hours of taking the previous dose. Samples collected during the first 4 days of dosing will be excluded from all analyses.

PK analyses will be performed on the Confirmed PK Sample Set from all everolimus patients in the Safety Set, which is defined as follows:

Confirmed PK Sample Set:

Cmin values collected prior to dose administration on the same treatment day and at 20-28 hours after previous dose, at steady state, and patient did not vomit within 4 hours of previous dose.

All analyses described below (except figures) will be repeated using a PK Sensitivity Sample Set if the Confirmed PK Sample Set contains less than 50% of the samples available among everolimus patients in the Safety Set.

PK Sensitivity Sample Set:

Cmin values collected prior to dose administration on the same treatment day and at 20-28 hours after previous dose.

Equivalent PK sample sets will be defined for the LTE Safety Set.

Biofluid concentrations will be expressed in mass per volume units. All missing data and concentrations below the lower limit of quantification (i.e., < 0.3 ng/ml) will be labeled as such in the concentration data listings, and will be excluded from all analyses.

3.13.1.2 PK definitions

Leading dose level

Leading dose level is the dose level that has the most immediate impact on the concentration, and for Cmin it is the dose of everolimus administered on the previous day.

Projected Cmin and time-normalized Cmin

The following model-based approach will be used to calculate the projected Cmin (instant Cmin). Cmin will be fitted using a modified power model expressed as a linear mixed effect model with a random effect for subject and fixed effects for each of up to the last 5 dosing

days (which corresponds to 120 hours, equivalent to 4 times everolimus half-life. Everolimus half-life is 30 hours according to the information in the Investigator Brochure). Cmin for subject i at timepoint j will be calculated as:

$$Cmin_{ij} = \mu + s_i + \sum_{k=1}^{5} \beta_k dose_{j-k} + \varepsilon_{ij}$$

Where

- μ is the population mean
- s_i is the random subject effect
- $dose_{j-k}$ is the dose k days prior to the C_{min} at time j
- ε_{ij} is the residual error
- Cmin and dose will be log-transformed and the age subgroup (stratification factor) will be included in the model as a covariate

The best fixed effects model will be selected based on the lowest Akaike Information Criterion (AIC) and clinical pharmacological interpretation, using the Maximum Likelihood (ML) method. Correlation between the 5 variables of dose will be assessed to determine appropriate possible model (to avoid collinearity issues). The final estimates will be obtained using restricted maximum likelihood (REML). Different covariate structures will be investigated. The final model will be used to predict Cmin for each subject at each day in the study based on the individual's dosing record. If the predicted C_{min} is negative, the projected Cmin will be set to zero.

Time-normalized (C_{min}, TN) is an estimate of the daily C_{min} for a patient i averaged over a time interval Δt =t2-t1 between consecutive assessment times (t1, t2). It will be calculated over each assessment interval (t_1, t_2) and will be linked to the response assessed (seizure frequency in this specific case) within the same interval in exposure-response analyses. For each patient, an overall $C_{min,TN}$ will be calculated across the maintenance period of the Core phase or other selected time intervals.

The time-normalized C_{min} will be defined as:

$$Cmin_{,TN} = AUC_{C(t1\text{-}t2)} \div (t_2 - t_1)$$

where $AUC_{C(t1\text{-}t2)}$ denotes the area under the pre-dose concentration-time curve over the interval (t_1, t_2), with the AUC calculated using the trapezoidal rule as follows

$$\left(\frac{(t_2 - t_1)}{2N}\right)\left(\sum_{k=1}^{k=N} pCmin(t_{k+1}) + pCmin(t_k)\right)$$

, where $pC_{min}(t_k)$ is the projected C_{min} at time $t_{k,,}$N= the number of equally spaces panels in the interval. N+1 = the number of grid points. t_1 = pCmin at the beginning of the interval. t_N = pCmin one day prior to t_2

If t1=t2, then $C_{min,TN} = pC_{min}(t1)$.

For patients with events occurring after treatment stop, projected Cmin after treatment stop will be computed from the last projected Cmin derived from the last dose (i.e. $C_{mini}(t_{last}+1)$,

where $t_{last}+1$ is the day following the last dose). Based on the PK linear properties of everolimus, ie, that the elimination rate does not depend on the concentration, the overall elimination constant, λ, can be approximated by the estimated half-life from population PK model. The computation of $C_{mini}(t_p)$ ($t_p > t_{last}+1$) will be as follows:

$$C_{mini}(t_p) = C_{mini}(t_{last}+1) \times \exp(-\lambda \times 24 \times (t_p - (t_{last} + 1)))$$

3.13.1.3 PK summary statistics

C_{min} will be summarized for each everolimus treatment arm by time window (as defined in Section 2.1.9.2). Descriptive statistics will include n, arithmetic mean, median, standard deviation, coefficient of variation, minimum and maximum. Geometric mean and geometric coefficient of variation will also be displayed (when ≥ 3 samples).

Cmin and dose-normalized Cmin (ng/mL per mg dose and ng/mL per mg/m^2 dose) will be summarized for each everolimus treatment arm by time window and by age subgroups (<6, 6 to <12, 12 to < 18, and ≥ 18 years of age) and by the use of CYP3A4 inducer.

Box plots of Cmin and dose-normalized Cmin (ng/mL per mg dose and ng/mL per mg/m^2 dose) will be plotted by time window and by age subgroups (<6, 6 to <12, 12 to < 18, and ≥ 18 years of age) (when ≥ 3 samples) and by the use of CYP3A4 inducer for each treatment arm.

Scatter plot of Cmin versus dose (mg) by age subgroups (<6, 6 to <12, 12 to < 18, and ≥ 18 years of age) (when ≥ 3 samples) will be plotted (use all Cmin values collected during the core and extension phase of the study).

Scatter plot of Cmin versus dose normalized to BSA (mg/m^2) by age subgroups (<6, 6 to <12, 12 to < 18, and ≥ 18 years of age) (when ≥ 3 samples) will be plotted (use all Cmin values collected during the core and extension phase of the study).

Scatter plot of dose-normalized Cmin (ng/mL per mg dose and ng/mL per mg/m^2 dose) versus age by the use of CYP3A4 inducer will be plotted (use all Cmin values collected during the core and extension phase of the study).

Time-normalized C_{min} calculated across the maintenance period of the Core phase will be summarized by treatment arm.

For the long-term evaluation of everolimus, these same analyses will be repeated for the patients originally randomized to either of the two everolimus arms. For placebo patients who start everolimus in the Extension phase, C_{min} will be summarized using different time windows that take account of the planned schedule of blood draws in the Extension phase, as defined in Section 2.1.9.2.

3.13.1.4 Impact of CYP3A4/PgP inducers and inhibitors

The impact of CYP3A4/PgP inducers and inhibitors on everolimus concentration by age range will be assessed by fitting a mixed model to log(C_{min}), with log(leading dose [mg/m^2]) as a fixed effect and patient as a random effect. The model will also include two indicator covariates and their interaction as fixed effects:

- co-administration of a CYP3A4/PgP inducer at the time of a sample (yes/no), defined as the intake of a CYP3A4 inducer or a PgP inducer over each of the previous 7 consecutive days
- co-administration of a CYP3A4/PgP inhibitor at the time of a sample (yes/no), defined as the intake of a CYP3A4 inhibitor or a PgP inhibitor on any one of the previous 4 days

Further covariates might be included if appropriate.

The estimate of the regression coefficients for the continuous covariates in the fitted model and their 90% confidence intervals will be presented. For the categorical factors, the geometric mean ratios and their 90% confidence intervals will be calculated. For the CYP3A4/PgP categories, the no inducer/no inhibitor category will be the reference. Note that including log-dose in the model is not done in an attempt to assess dose-proportionality.

3.13.1.5 Exposure-response relationship

The relationship between everolimus exposure and the two primary efficacy endpoints of response rate and percentage change from baseline in seizure frequency will be investigated.

For response rate, logistic regression will be used to model the probability of response. The model will include terms for time-normalized Cmin (log-transformed) in the maintenance period of the Core phase (defined as from study day 43 until the last day of study medication in the Core phase and extension phase), baseline seizure frequency (SF_B defined in Section 3.9.1.1), and will be stratified by age subgroup. Further covariates might be included if appropriate.

To estimate the exposure-response relationship on the seizure frequency, a multiplicative linear regression model will be used. The dependent variable will be the absolute seizure frequency during the maintenance period of the Core phase and extension phase (in log scale). The model will include baseline seizure frequency (SF_B) (in log scale) and time-normalized Cmin (in log scale) in the maintenance period of the Core phase as fixed effects. Further covariates might be included if appropriate. The absolute seizure frequency change for a 2-fold increase in Cmin will be also estimated. Box plots of percentage reduction from baseline in seizure frequency by range of time-normalized everolimus concentration at through (Cmin,TN) in steps of 1ng/mL, during the maintenance period will also be presented.

An additional analysis will be performed considering the seizure frequency (log scale) and the time-normalized Cmin for fixed time intervals (e.g. 2 weeks), accounting for the longitudinal nature of the data and the inter-patient variability. A linear mixed model with time-normalized Cmin, baseline seizure frequency, both log transformed as fixed effects and patient as random effect will be performed. Further covariates might be included if appropriate. In particular, CYP3A4/PgP inducers (yes/no) and use of CYP3A4/PgP inhibitors (yes/no) will be considered. Scatter plots of percentage change from baseline in seizure frequency versus time-normalized Cmin with a line joining data of individual patients during the maintenance period of the Core phase and extension phase will be presented, separately for different values of baseline seizure frequency (in units of seizures per week: < 3, 3 to < 6, 6 to < 9, ≥ 9). Note that if there are insufficient numbers of patients in any of the groups these categories of seizure frequency might be changed. Targeted concentrations ranges of 3 to 7 ng/mL and 9 to 15 ng/mL will be identified by lines on the graph.

Response rate (as defined in Section 3.9.1.1) and percentage reduction in seizure frequency (as defined in Section 3.9.1.2) will be analyzed by time normalized Cmin across the maintenance phase categories (< 3, 3 to 7, > 7 to < 9, 9 to 15 and > 15 ng/ml). Response rates will be provided with exact 95% confidence intervals (see Section 4.4) for each category. The median percentage reduction from baseline will be presented for each treatment group along with 95% bootstrap confidence intervals (see Section 4.7) for each category.

Time-normalized Cmin values associated with ≥ 50% reduction from baseline in seizure rate (Cmin_responder) will be estimated for individual patients. Frequency distribution curve of Cmin_responder (10, 25, 50, 75, 90 percentiles of Cmin associated with ≥ 50% reduction from baseline in seizure rate) will be made.

Time-normalized Cmin values associated with 40% to 60% reduction from baseline in seizure rate (Cmin_50%) will be estimated for individual patients. Frequency distribution curve of Cmin_50% (10, 25, 50, 75, 90 percentiles of Cmin associated with ~ 50% reduction from baseline in seizure rate) will be made.

Exposure-response analyses will consider Core phase data. In addition, exposure-response analyses will be repeated considering all data available at the time of the cut-off, including data beyond the core phase.

3.13.1.6 Exposure-safety relationship

Selected summary AE tables will be presented split by exposure to everolimus, in terms of time-normalized C_{min} over the entire Core phase (titration period plus maintenance period), in order to compare the safety profiles. The selected AE will be the clinically notable AE groupings, as defined in Section 3.10.1.6. The tables will present the number of patients experiencing a given AE. Time-normalized C_{min} values will be categorized in units of ng/ml as < 3, 3 to 7, > 7 to < 9, 9 to 15 and > 15.

In addition the time to first event of selected AEs will be fitted using an extended Cox model with interval time-normalized C_{min} as a time varying covariate, stratified by age subgroups. Four time intervals will be considered to adjust relative risks by the impact of exposure on safety: during the 6 weeks titration period, the 12 weeks of the maintenance period of the core phase, the 8 weeks titration period of the extension phase and the extension phase (after week 26). Relative risks of the AEs for a two-fold increase in exposure and its 90% confidence interval will be reported.

Exposure-safety analyses will be based on available data at the time of the cut-off, that means even data points beyond the core phase will be also considered.

3.13.2 AED PK

The concentration levels of 12 commonly prescribed AEDs that are CYP3A4/PgP inducers or inhibitors will be assessed, in order to investigate the possibility of drug-drug interactions with everolimus. The 12 AEDs are of interest consist of five CYP3A4/PgP inducers: carbamazepine, oxcarbamazepine, phenobarbital, phenytoin and primidone, and seven CYP3A4/PgP inhibitors: clonazepam, diazepam, clobazam, felbamate, topiramate, valproic acid and zonisamide.

All patients should be taking a background regimen of one to three AEDs at a stable dose from 4 weeks prior to screening through until the end of the Core phase. Blood samples for AED PK are scheduled to be taken at two time points in the Baseline phase, at the screening visit (week -8) and the baseline visit (day -1), providing two values of AED concentration in the absence of everolimus. Two further samples will be taken at weeks 1 and 3 of the Core phase, providing two values of AED concentration in the presence of everolimus. Samples will only be collected from patients who are taking at least one of the 12 AEDs of interest, and who are over 20 kg in body weight (to avoid drawing too much blood in the very youngest patients, given that blood draws are also required for laboratory determinations and everolimus PK). If a patient is taking more than one of the AEDs of interest in their background AED regimen, then a separate sample will be obtained for each AED.

Blood samples for AEDs must be pre-dose, obtained prior to dose administration on the same treatment day.

Concentrations of each AED will be summarized for each everolimus treatment arm by time window (as defined in Section 2.1.9.3). Descriptive statistics will include n, arithmetic mean, median, standard deviation, coefficient of variation, minimum and maximum. Geometric mean and geometric coefficient of variation will also be displayed (when ≥ 3 samples).

The impact of everolimus on the exposure of the AEDs will be assessed via linear mixed models to compare the exposure of the AEDs before and after the administration of everolimus. Separately for each AED, a linear mixed model will be fitted to the log-transformed concentrations, and will include period (before and after everolimus administration) as a fixed effect, and patient as a random effect. Geometric mean ratios of the concentrations with and without everolimus (as reference) and their 90% confidence intervals will be calculated from the model.

The above analyses will be repeated considering only patients exposed to only one of the 12 AEDs of interest to investigate any potential confounding effect.

Model-based analysis will be performed only on those AEDs taken by a minimum of 6 patients with valid concentrations for both everolimus and the corresponding AED.

3.14 Subgroup analyses

Subgroup analyses will be performed to compare each everolimus arm versus placebo as detailed in Table 3-6 below.

Table 3-6 Subgroup analyses

Subgroup	Categories	Reference	Use
Age	Calculated from date of birth and randomization date as collected on the eCRF, as follows: 1 to <6 years, 6 to <12 years, 12 to <18 years and ≥ 18 years	Stratification factor	(1) Safety (2) Efficacy
Gender	Male versus Female	Regulatory requirement	(1) Safety (2) Efficacy

Subgroup	Categories	Reference	Use
Race	As collected in eCRF if large enough to allow interpretable results. If small, grouping of categories might be considered.	Regulatory requirement	(1) Safety (2) Efficacy
Ethnicity	As collected in eCRF if large enough to allow interpretable results. If small, grouping of categories might be considered.	Regulatory requirement	(1) Safety (2) Efficacy
[redacted]			
Japanese*	Japanese	Regulatory requirement (PMDA)	(1) Safety (2) Efficacy

(1) Main safety analysis (overall AE tables, adverse events of special interest) will be repeated.
(2) If the primary analysis of response rate or percentage reduction in seizure frequency is statistically significant, subgroup analyses will be performed.

* Subgroup analyses of the main efficacy and safety endpoints, as well as baseline characteristics, will be performed on patients randomized in Japan, for use by the Japanese regulatory authority.

[redacted] No adjustment of alpha for multiplicity or any other strategy to control the family-wise error rate will be implemented.

3.15 Interim analyses

No interim analysis for efficacy is planned, but there will be DMC reviews of the ongoing safety data as the trial progresses. The first meeting will review data from the first 6 months of the study, with subsequent meetings every 6 months thereafter until the study is unblinded.

As outlined in the DMC charter, the safety review outputs will be prepared by an independent statistician and an independent programmer from Novartis, neither of whom will belong to the trial team or will be involved in any other aspects of the trial conduct. Semi-blinded results will be shared with the DMC members using a secured web portal to which Novartis personnel have no access (except the independent statistician/programmer). After the meeting, the DMC chair will provide a recommendation to the Novartis Oncology global development head. Apart from the recommendation made, the Novartis team members will remain blinded to any study results.

Details of all outputs to be provided to the DMC will be described in a separate DMC RAP document.

3.16 Sample size calculation

The sample size was chosen to provide adequate power for the primary objective comparing seizure frequency between each everolimus arm and the placebo arm. The sample size calculation provided here is based exclusively on response rate, the EMA primary endpoint, but it is also expected to provide sufficient patients for the power of the FDA primary endpoint, percentage reduction in seizure frequency. This is because response rate is a binary transformation of percentage reduction in seizure frequency, and therefore likely to be less sensitive owing to the loss of information going from a continuous variable to binary.

It was assumed that response rates would be 15% in the placebo arm and 35% in each of the two everolimus arms. That is, there was no *a priori* strong expectation that the higher targeted trough everolimus arm 9 to 15 ng/mL would deliver a higher response rate than the lower targeted trough everolimus arm 3 to 7 ng/mL, as better efficacy may be mitigated by worse tolerability. For this reason, the testing strategy was to simultaneously compare each pairwise comparison, splitting the significance level, rather than testing hierarchically starting with the higher trough arm for example.

Using nQuery version 6.1 it was determined that a sample size of 355 patients would ensure 90% power for each of the primary comparisons of each everolimus arm versus placebo, assuming one-sided 1.25% significance levels for each CMH chi-square test, and assuming balanced randomization (i.e., 115 patients per randomization arm).

Table 3-7 Sample size and power for primary objective

Comparison	One-sided alpha	Response rates under alternative hypothesis	Sample Size	Power
Everolimus 3-7 ng/mL arm versus Placebo arm	1.25%	35% vs. 15%	230 (115 vs. 115)	90%
Everolimus 9-15 ng/mL arm versus Placebo arm	1.25%	35% vs. 15%	230 (115 vs. 115)	90%

Approximately 5 months after the first patient was randomized, a human error was identified which caused the IRT system not to perform dose titrations despite Cmin values outside the targeted trough range. The IRT system was updated the same day. By that time a total of 47 patients had already passed their Week 2 assessment, the point at which the first dose titration could have been made.

After an investigation into the potential impact of this error on the study power, it was noted that placebo patients would be unaffected, and in all likelihood, the majority of patients in the everolimus 3 to 7 ng/ml arm would be expected to achieve everolimus concentrations within the targeted range. However, since few patients randomized to the everolimus 9 to 15 ng/ml arm would be expected to achieve adequate dosing, it was decided to increase the sample size in the everolimus 9 to 15 ng/ml arm by 10 patients.

The choice of 10 patients was made on the following grounds. Based on the pattern of complete and incomplete blocks, up to 18 of the 47 potentially affected patients may have been on the everolimus 9 to 15 ng/ml arm. As a worst case, assume all 18 patients were under-

dosed and achieved everolimus concentrations in the 3 to 7 ng/ml range, and assume that this trough range is no better than placebo. Then at the end of the study, among the 115 planned patients on the everolimus 9 to 15 ng/ml arm, 18 patients will have a 15% chance of being responders (i.e., as placebo arm patients) and 97 will have a 35% chance. Overall this makes an expected response rate in the everolimus 9 to 15 ng/ml arm of 31.9%, in which case the power for the comparison with the placebo arm drops to 78%. Other more realistic scenarios lead to a smaller loss in power, for example with 12 patients under-dosed and subject to placebo arm response rate, the power is 83%, with 6 patients it is 87%. The sample size increase was chosen to protect against loss in power for these more realistic scenarios, where the addition of 10 patients to the everolimus 9 to 15 ng/ml arm improves power to 85% for 12 under-dosed patients and to 88% for 6 under-dosed patients.

The sample size increase of 10 patients was made by inserting a number of blocks with randomization ratio of 1:1:2 in favor of the everolimus 9 to 15 ng/ml arm, with the planned sample size becoming 355 patients (115 on placebo arm, 115 patients on everolimus 3 to 7 ng/ml arm and 125 patients on everolimus 9 to 15 ng/ml arm; overall randomization ratio of 1:1:1.09). The required modification of the randomization list was overseen by an Independent Randomization Expert, a Novartis employee who is not part of the study team.

4 Details of the statistical analysis

4.1 Baseline comparability

Baseline variables refer to characteristics of patients at the start of a study (see Section 2.1.7 for more detailed definition of ‘baseline’).

Baseline variables will be descriptively summarized by treatment group. No inferential analysis will be performed to compare baseline characteristics between the treatment arms.

4.2 One-sided vs. two-sided test

Since this is a placebo controlled study where the primary objective is clearly directional, the hypothesis test for the primary endpoint will be one-sided at the 2.5% level of significance.

However, confidence intervals will always be two-sided and at the 95% level, since they will be used for estimation rather than decision-making.

4.3 Comparison of response rates

Cochran-Mantel-Haenszel test

The Cochran-Mantel-Haenszel (CMH) chi-square test (implemented via SAS procedure FREQ with the CMH option in the TABLES statement) will be used to test the difference in response rates between the treatment arms. The p-value will be obtained from the “General association” CMH statistic.

The one-sided p-value will be obtained using the square root of the chi-square distributed CMH general association statistic. As a consequence, the one-sided p-value will then be either

the two-sided p-value divided by two or its opposite (1 – two-sided p-value divided by two) depending on the direction of the Odds ratio estimate.

The following SAS code will be used:

```
PROC FREQ data=dataset;
TABLES stratum*trt*response / CMH;
OUTPUT OUT=pval_cmh COMOR CMHGA;
RUN;
/* stratum represents stratum variable
   trt represents treatment group variable
   response represents response variable   */

DATA pval_cmh1sid;
      SET pval_cmh;
      IF _MHOR_>1 THEN pvalue=P_CMHGA/2;
       ELSE pvalue=1-P_CMHGA/2;
run;
* If OR is in favor of everolimus (i.e. <1) then one-sided p-
value = two-sided p-value/2 else one-sided p-value = 1- two-
sided p-value/2;
/* pvalue represents the one-sided p-value
   P_CMHGA represents the two-sided p-value from the "General association"
   CMH statistic
   MHOR_represents the adjusted Mantel-Haenszel odds ratio*/
```

Note: The direction of the odds ratio depends on the parametrization of the model. In the above example, it is assumed that the probability modeled is absence of response and the reference arm is placebo. In other words, an odds ratio below one indicates that the probability of being non-responder is less in the everolimus arm compared to placebo.

Odds ratio

The odds ratio can also be used as a measure of association between treatment and response. It will be derived from a logistic regression model (implemented using SAS procedure LOGISTIC, with treatment specified as an explanatory variable in the CLASS statement) which allows for including not only the stratification factor but also for adjustments for other covariates, both categorical and continuous. The odds ratio will be presented with 95% Wald confidence limits.

The following SAS code will be used:

```
PROC LOGISTIC data=dataset;
     CLASS trt;
      MODEL response=trt stratum <covariate_1>..<covariate_k>;
RUN;
/* stratum represents stratum variable
   trt represents treatment group variable
   response represents response variable
   covariate_i (i=1,…,k) represents the ith covariate */
```

4.4 Confidence interval for response rate

An exact binomial confidence interval (implemented using SAS procedure FREQ with the EXACT statement for one-way tables) will be used (Clopper and Pearson 1934).

4.5 Rank Analysis of Covariance

Continuous data where the normality assumption is unlikely to be valid will be analyzed using rank ANCOVA (rank analysis of covariance). Rank ANCOVA is a nonparametric approach and is preferred to classical parametric ANCOVA when the data may not be even approximately normally distributed. This method does not lead to much loss of power versus parametric ANCOVA should the data turn out to be approximately normal (Canover 1999).

Rank ANCOVA will be implemented in SAS using the procedures RANK, REG and FREQ. The following three steps illustrate how to compare a continuous response variable (labeled *perc*) between two treatment groups (labeled *treat*), adjusting for a covariate (labeled *cov*), and with an analysis stratified by a stratification factor (labeled *strat*).

1. Procedure RANK is used to calculate fractional ranks for the covariate and the response variable across the combined treatment groups separately within each stratum. Fractional ranks are obtained by dividing each rank by the number of patients with non-missing data in the stratum; fractional ranks are used because they adjust for possibly different numbers of patients in each stratum. In case of ties, the mean of the ranks will be assigned.

```
PROC RANK data=dataset FRACTION TIES=MEAN OUT=ranks;
BY strat;
VAR cov perc;
RUN;
```

2. Procedure REG is then used to fit separate linear regression models for each stratum. In each model the fractional ranks of the response variable is the dependent variable, and the fractional ranks of the covariate is the independent variable. The residuals from each model are stored in an output dataset.

```
PROC REG data=dataset NOPRINT;
BY stratum;
MODEL perc=cov,
OUTPUT OUT=residual R=resid;
RUN;
```

3. Then a stratified Cochran-Mantel-Haenszel mean score test is performed using the residuals as scores, and a p-value for the treatment comparison is obtained.

```
PROC FREQ data=residual;
TABLES stratum*treat*resid / NOPRINT CMH2;
OUTPUT OUT=pval_rank CMH2;
RUN;
```

The p-value is taken from the SAS output labeled “Row Mean Scores Differ”.

The one-sided p-value will be obtained using the square root of the chi-square distributed CMH row mean score statistic. As a consequence, the one-sided p-value will then be either the two-sided p-value divided by two or its opposite (1 – two-sided p-value divided by two) depending on the sign of the mean difference estimate of the percentage change from baseline.

The following SAS code will be used:

```
ODS OUTPUT DIFFS=diffs LSMEANS=lsmeans;
PROC MIXED data=residual;
     CLASS strata trtreg1c;
     MODEL resid=strata trtreg1c /ddfm=satterth;
     LSMEANS trtreg1c /diff=control("Placebo");
run;

DATA pval_rank1sid;
      MERGE pval_rank diffs;
      IF Estimate>0 THEN pvalue=P_CMHRMS/2;
       ELSE pvalue=1-P_CMHRMS/2;
run;
* If the mean percentage reduction is in favor of everolimus
(i.e. the mean percentage reduction is higher in everolimus)
then one-sided p-value = two-sided p-value/2 else one-sided p-
value = 1- two-sided p-value/2;
/* pvalue represents the one-sided p-value
   P_CMHRMS represents the two-sided p-value from the "Row Mean Scores" CMH
   statistic
   Estimate represent the mean difference between everolimus and placebo*/
```

4.6 Bonferroni-Holm method

The Bonferroni-Holm method for multiple testing (Holm 1979) was designed to control the family-wise error rate, that is, the probability of incorrectly rejecting at least one of a family of hypotheses. It is more powerful than the Bonferroni approach, and does not require any assumptions (e.g., distributional, independence), and can therefore be applied to any family of pairwise comparisons regardless of the joint distribution of the test statistics. It is a conservative test and the family-wise error rate cannot exceed the planned alpha level.

The method works as follows.

Consider testing the family of hypotheses H_{0i}, where $i = 1, 2, \ldots, k$, with the intention of controlling the family-wise error rate to α (e.g., $\alpha = 0.05$).

Let p_i, $i = 1, 2, \ldots, k$, denote the sample p-values of tests for H_{0i}, $i = 1, 2, \ldots, k$, where the p-values were computed without any multiplicity adjustment.

Let [1], [2], …, [k] denote indices such that $p_{[1]} \leq p_{[2]} \leq \ldots \leq p_{[k]}$.

That is, $p_{[1]}$ refers to the smallest p-value, and in general, [i] is the inverse rank of p_i among p_1, p_2, …, p_k.

The Bonferroni-Holm method then proceeds as follows.

Step 1: If $p_{[1]} \leq \alpha \div k$, reject $H_{0[1]}$ and go to Step 2; otherwise stop.

Step 2: If $p_{[2]} \leq \alpha \div (k - 1)$, reject $H_{0[2]}$ and go to Step 3; otherwise stop.

…

Step k: If $p_{[k]} \leq \alpha$, reject $H_{0[k]}$ and stop.

There is no direct procedure in SAS to run the Bonferroni-Holm method, so it will be run by reading in the appropriate unadjusted p-values and the required alpha level, and then performing the steps 1 though k above.

4.7 Bootstrap method

To calculate the confidence intervals of observed medians, percentile bootstrap confidence intervals will be used (Efron and Tibshirani 1993).

From the original sample of size n, a total of B independent bootstrap samples are selected, each consisting of n data values drawn with replacement. The number of bootstrap samples will be B = 100 000.

The sample median is evaluated for each of the B independent bootstrap samples.

The B sample medians are ordered to obtain the 2.5th percentile and 97.5th percentile.

The 95% confidence interval itself will then be: (2.5th percentile, 97.5th percentile).

Bootstrap methodology (Efron and Tibshirani 1993) stratified by age subgroup will be used to estimate the difference between the median percentage reduction from baseline between each of everolimus treatment arms and placebo using the following steps:

1. Let N1j, N2j, N3j denote the numbers of patients in the relevant analysis population who were randomized to the everolimus 3 to 7 ng/ml arm (Arm 1), the everolimus 9 to 15 ng/ml arm (Arm 2) and the placebo arm (Arm 3) respectively and stratum j, j=1 (1 to <6 years), 2 (6 to <12 years), 3 (12 to <18 years) and 4 (≥ 18 years). Calculate the median percentage reduction from baseline in seizure frequency in the 3 arms and by stratum M1j, M2j and M3j (M=median). Calculate the difference between the median percentage reduction from baseline between each of everolimus treatment arms and placebo: Y1j as M1j-M3j and Y2j as M2j-M3j.

2. Take a sample with replacement, of size N1j from patients in Arm 1 and stratum j, of size N2j from patients in Arm 2 and stratum j, and of size N3j from patients in Arm 3 and stratum j. Calculate an estimate of Y1j and Y2j. Repeat this resampling estimation step Nsim=10,000 times. Denote the estimates of Y1j and Y2j obtained from these steps by Y1ij and Y2ij, i=1 to Nsim.

3. Calculate the variance of the estimates Y1ij and Y2ij by stratum and denote it Var(Y1j*) and Var(Y2j*) respectively. Calculate the weight of each stratum equal to W1j=[1/Var(Y1j*)] / [Σj=1 to j=4 1/Var(Y1j*)] for the comparison between the everolimus 3 to 7 ng/ml arm and the placebo arm, and W2j=[1/Var(Y2j*)] / [Σj=1 to j=4 1/Var(Y2j*)] for the comparison between the everolimus 9 to 15 ng/ml arm and the placebo arm.

4. Calculate for each sample, the stratified estimates Y1iW=Σj=1 to j=4 Y1ij*W1j and Y2iW=Σj=1 to j=4 Y2ij*W2j

5. Calculate the stratified estimate of comparison as Y1=Σj=1 to j=4 Y1j*W1j for everolimus 3 to 7 ng/ml versus placebo and Y2=Σj=1 to j=4 Y2j*W2j for everolimus 9 to 15 ng/ml versus placebo.

6. To obtain the 95% confidence interval, Y1 and Y2 are ordered to obtain the 2.5th percentile and 97.5th percentile.

4.8 Time-to-event analyses

This section presents the general methodology used to analyze time-to-event variables.

4.8.1 Kaplan-Meier estimates

An estimate of the time-to-event function in each treatment group will be constructed using the Kaplan-Meier (product-limit) method as implemented in the procedure LIFETEST with the METHOD=KM option. Kaplan-Meier curves will be displayed by treatment group.

Median time-to-event for each treatment group will be obtained along with 95% confidence intervals using the method of Brookmeyer and Crowley 1982.

Kaplan-Meier estimates of the time-to-event function with 95% confidence intervals at 6 and 12 months, where 1 month is defined as (365.25/12)=30.4375 days, will be summarized by treatment group. Confidence intervals will be constructed using Greenwood's formula for the standard error of the Kaplan-Meier estimate (Collett 1994, p.23). The log-log transformation of the survivor function (Kalbfleisch and Prentice 1980) will be used in order to ensure that the confidence limits remain in the interval [0, 1]. The log-log transformation is implemented in the LIFETEST procedure using the option CONFTYPE=LOGLOG.

Kaplan-Meier graphs will be constructed using S-Plus software, although the statistics displayed on the graph (e.g., median, hazard ratio) will be obtained from SAS.

4.8.2 Hazard ratio

Hazard ratio as a treatment effect measure will be derived from the Cox proportional hazards model using the SAS procedure PHREG (with the TIES=EXACT option in the MODEL statement).

A stratified Cox model will be used (where the baseline hazard function is allowed to vary across strata). In this model the MODEL statement will include the treatment group variable as the only covariate and the STRATA statement will include the protocol stratification variable (age subgroup). The hazard ratio and its two-sided 95% confidence interval will be based on the Wald test.

4.9 Comparison of means

The mean difference along with its 95% CI between two continuous variables will be calculated using a t-test assuming equal variances in the two arms. The mean difference (95%CI) will be obtained from the "Diff (1-2) pooled method" estimate.

The following SAS code will be used:

```
PROC TTEST data=dataset TEST=DIFF;
      CLASS trt;
      VAR response;
run;
/* trt represents treatment group variable
```

```
        response represents response variable   */
```

4.10 Linear Mixed model for Expose-Response relationship

A linear mixed model analysis will be performed to investigate the relationship between the frequency of seizures and the time-normalized Cmin assessed every two weeks. The linear mixed model allows introducing fixed as well as random effects and therefore adjusting for the repeated measures nature of the data. Variables will be log transformed. Different covariance patterns will be investigated. The best model will be selected based on the lowest Akaike Information Criteria (AIC, BIC or AICC) and clinical pharmacology interpretation, under maximum likelihood (ML) method of estimation. Once the fixed effect model is selected, the final estimates will be obtained based on a restricted ML (REML) method.

Marginal estimates and corresponding 95% confidence interval will be presented for each factor in the model. Interactions among the different covariates will be investigated and kept in the model if the statistical significance is below the 10%.

PROC MIXED procedure in SAS could be used.

```
      PROC MIXED data=dataset   method=(ML or REML);
      Class patient cov1 cov2 ;
      model Y =  logCmin cov1 cov2 / DDFM=  ;
      random intercept/subject=;
RUN;
```

The estimate of the 2 fold increases in Cmin will be calculated as follows: fold2Est=**100***(exp((log(Estimate)*log(**2**)) - **1**)) where the 'Estimate' is the one given by the model. The same formula will be applied for the 95% CI.

For simplicity the same procedure will be used to fit a linear regression model to investigate the relationship between frequency of seizures at the end of the core phase and the time-normalized Cmin over the entire core phase period.

4.11 Stratified Cox model and Extended Cox model for Exposure-Safety relationship

Hazard ratio will be derived from the ***Cox proportional hazards model*** using SAS procedure PHREG (with TIES=EXACT option in the MODEL statement).

Hazard ratio with two-sided 95% confidence interval will be based on ***Wald test***.

The Stratified Cox model adjusted for covariates will be used. Subgroup of age and other covariates as appropriate will be considered.

```
PROC PHREG data=dataset;
      MODEL survtime*censor(1)=cov1..<covk>;/TIES=EXACT;
      STRATA stratum1 .. <stratum k>;
RUN;
/* survtime represents variable containing event/censor times;
   censor represents censoring variable (1=censored, 0=event);
```

```
   cov1 to covk represent covariates;
   STRATA statement to be used for stratified studies only,
   stratum1 to stratumk represent stratification variables */
```

The Extended Cox model in those cases where a time varying covariate will be included. Subgroup of age and other covariates as appropriate will be considered.

```
PROC PHREG data=dataset;
      MODEL (start stop)*censor(1)= cov1..<covk>;/TIES=EXACT;
      STRATA stratum1 .. <stratum k>;
RUN;
/* censor represents censoring variable (1=censored, 0=event);
   start and stop refers to the lower and upper limit of the given time interval.
The lower limit of the interval is not included while the upper it is. For a
patient with no event, the start and stop will refer to the beginning and end of
the interval. For a patient with an event, the start refers to the beginning of
the interval.
   cov1 to covk represent covariates;
   STRATA statement to be used for stratified studies only,
   stratum1 to stratumk represent stratification variables */
```

5 References

Baumgartner RN, Roche AF, Himes (1986). Incremental growth tables: supplementary to previously published charts. American Journal of Clinical Nutrition, 43, 711-722.

Brookmeyer R and Crowley J (1982). A Confidence Interval for the Median Survival Time, Biometrics, 38, 29-41.

Canover WJ (1999). Practical Nonparametric Statistics (Third Edition), Wiley, New York.

Clopper CJ, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26, 404-413.

Collett D (1994). Modelling survival data in medical research. Chapman & Hall, London.

Cramer JA. Van Hammee G. N132 Study Group. (2003) Maintenance of improvement in health-related quality of life during long-term treatment with levetiracetam. Epilepsy & Behavior, 4, 118-123.

Cramer, J. A., Westbrook, L. E., Devinsky, O., Perrine, K., Glassman, M. B. and Camfield, C. (1999), Development of the Quality of Life in Epilepsy Inventory for Adolescents: The QOLIE-AD-48. Epilepsia, 40, 1114–1121.

DuBois D and DuBois EF (1916). A formula to estimate the approximate surface area if height and weight be known. Arch Intern Medicine, 17, 863-71.

Efron B and Tibshirani RJ (1993). An Introduction to the Bootstrap. Chapman & Hall, London.

Fenichel P (2012). Delayed puberty. Endocr Dev, 22, 138-59.

Holm, S. (1979). A simple sequential rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65-70.

International Conference on Harmonization. E9: Statistical Principles for Clinical Trials.

Kalbfleisch JD and Prentice RL (1980). The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons, Inc.

QOLIE Development Group (2013) Scoring Manual for the QOLIE-31-P: Patient-weighted quality of life in epilepsy.

Sabaz M, Lawson AJ, Cairns RD, Duchowny MS, Resnick TJ, Dean PM, Bye AME (2003). Validation of the Quality of Life in Childhood Epilepsy Questionnaire in American epilepsy patients. Epilepsy and Behaviour, 4(6), 680-691.

Sparrow SS, Cicchetti DV, Balla DA (2005). Vineland-II Survey Forms Manual. Second Edition, PsychCorp, Bloomington, MN.

Stokes ME, Davis CS, Koch GG (2000). Categorical data analysis using the SAS system. Second Edition, Cary, NC, SAS Institute Inc.

Tanner JM, Goldstein H, Whitehouse RH (1970). Standards for children's height at ages 2-9 years allowing for height of parents. Archives of Disease in Childhood, 45, 755-762.

Wechsler D, Naglieri JA (2006). Wechsler Nonverbal Scale of Ability Administration and Scoring Manual, Pearson, Bloomington, MN.

NOVARTIS

Clinical Development

Afinitor®/Votubia®/Everolimus

CRAD001M2304 / NCT02962414

A three-arm, randomized, double-blind, placebo-controlled study of the efficacy and safety of two trough-ranges of everolimus as adjunctive therapy in patients with tuberous sclerosis complex (TSC) who have refractory partial-onset seizures

Statistical Analysis Plan (SAP) – Addendum1
to support the Clinical Study Report (CSR) of the Extension phase (02-Sep-2016 cutoff date)

Author: [redacted]

Document type: SAP Documentation

Document status: Final version

Release date: 10-Feb-2017

Number of pages: 11

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
18-Oct-2016	Prior to DB lock	Creation of final version	N/A - First version	NA
09-Feb-2017	After DBL	[redacted]	Addendum 1	Sections 2.1.5.2, 2.1.7, 2.1.8 and 3.1.2 added

Table of contents

List of abbreviations

AED	Antiepileptic drug
CSR	Clinical Study report
EEG	Electroencephalogram
FSH	Follicle Stimulating Hormone
LH	Luteinizing hormone
LTE	Long-Term Evaluation
SAP	Statistical Analysis Plan
TSC	Tuberous Sclerosis Complex

1 Introduction

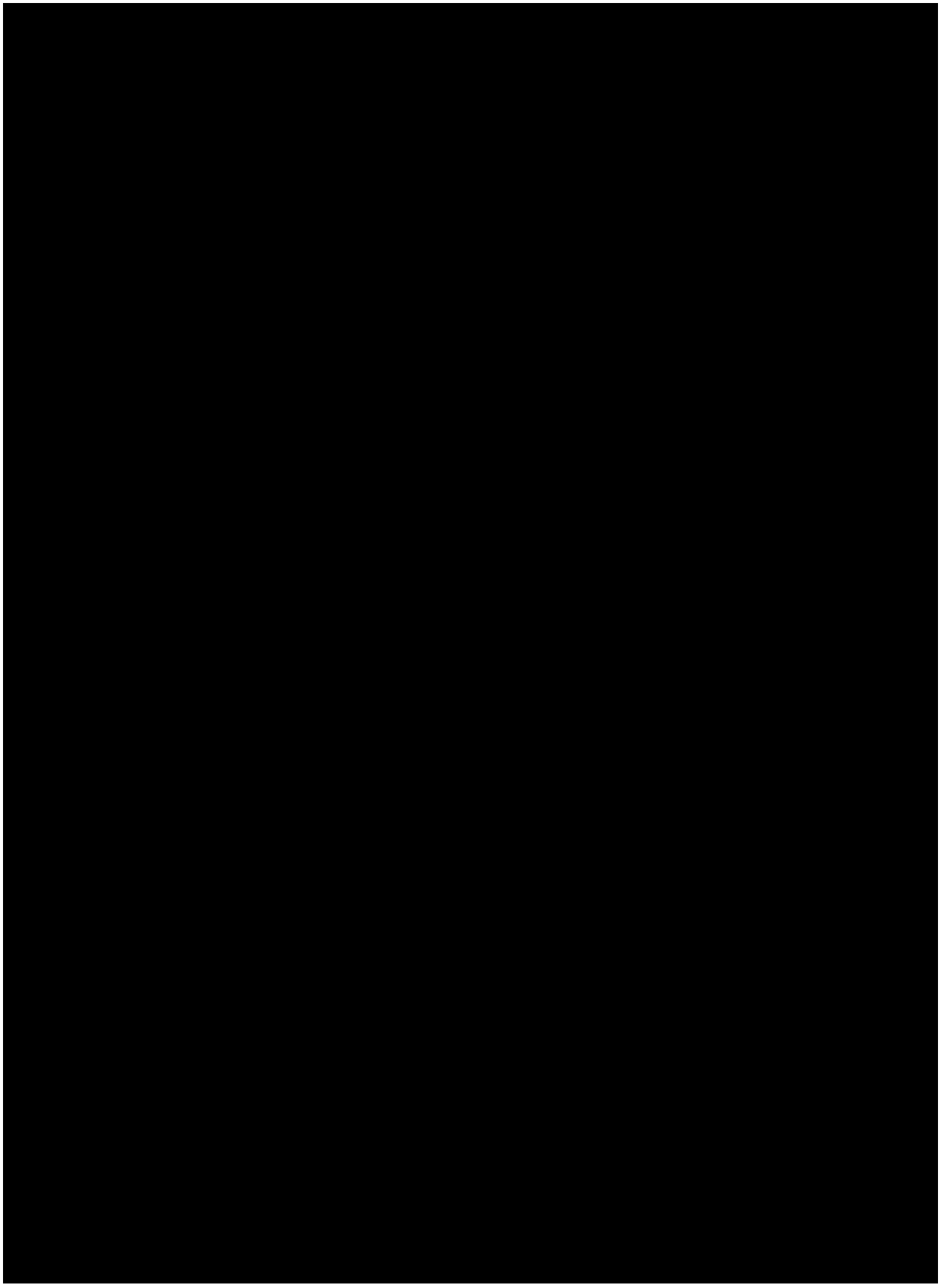
Analyses conducted at the end of the core phase, focusing on the Long-Term Evaluation (LTE) of everolimus using all available data from both the Core and Extension phases of the trial were described in the Statistical Analysis Plan (SAP) "CRAD001M2304_M3: Detailed statistical methodology_Final analysis_Addendum 1".

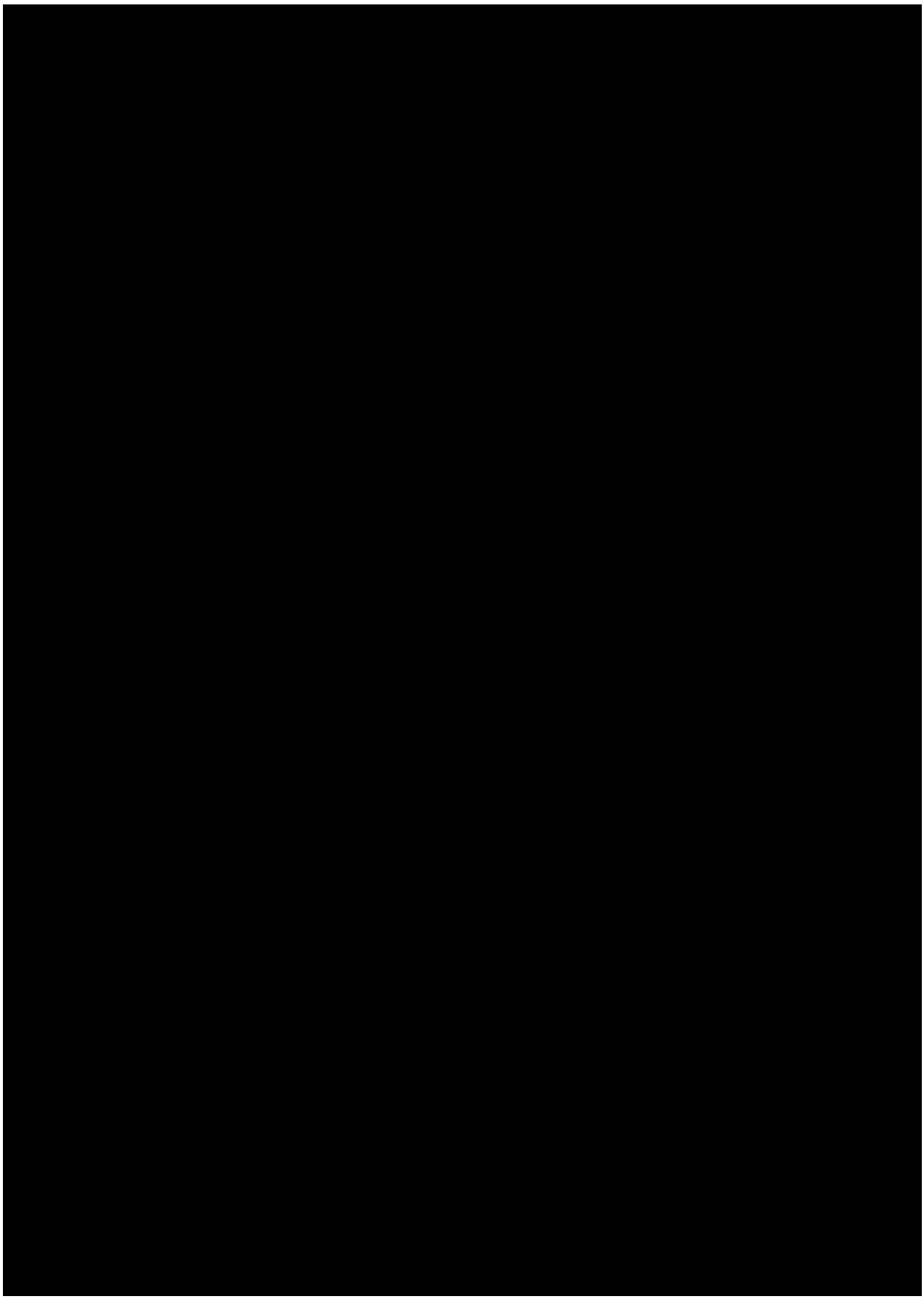
The purpose of this new ('SAP Extension CSR') is to describe the additional exploratory analyses planned to further assess the LTE of everolimus in this study population (TSC with refractory seizures). Both SAPs will support the writing of the Clinical Study Report (CSR) of the Extension phase. The cutoff date for these analyses will be on 02-Sep-2016, at the end of the extension phase, which is 48 weeks after the last patient completed the core phase.

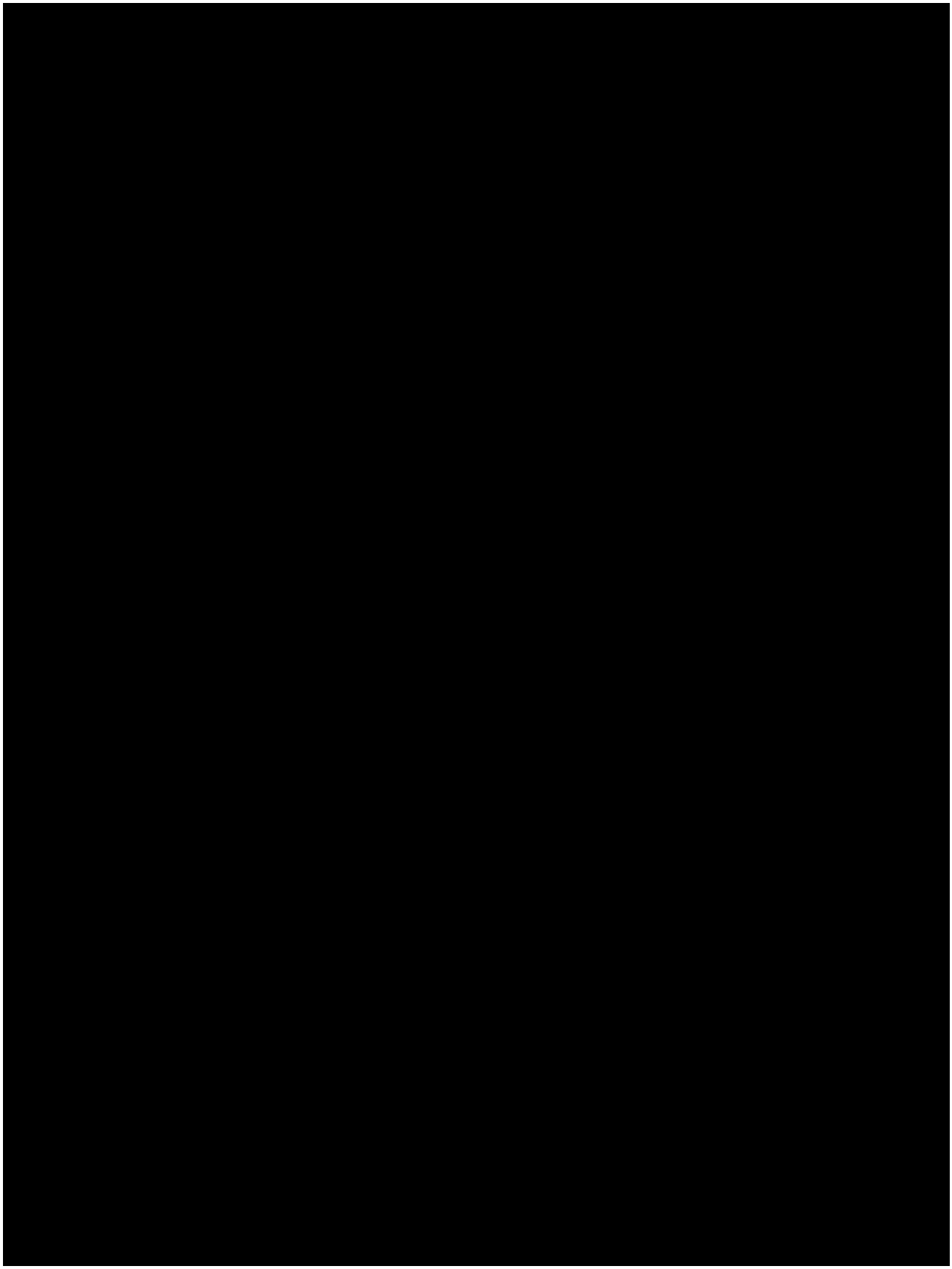
The CSR of the Extension phase will include the following analyses:

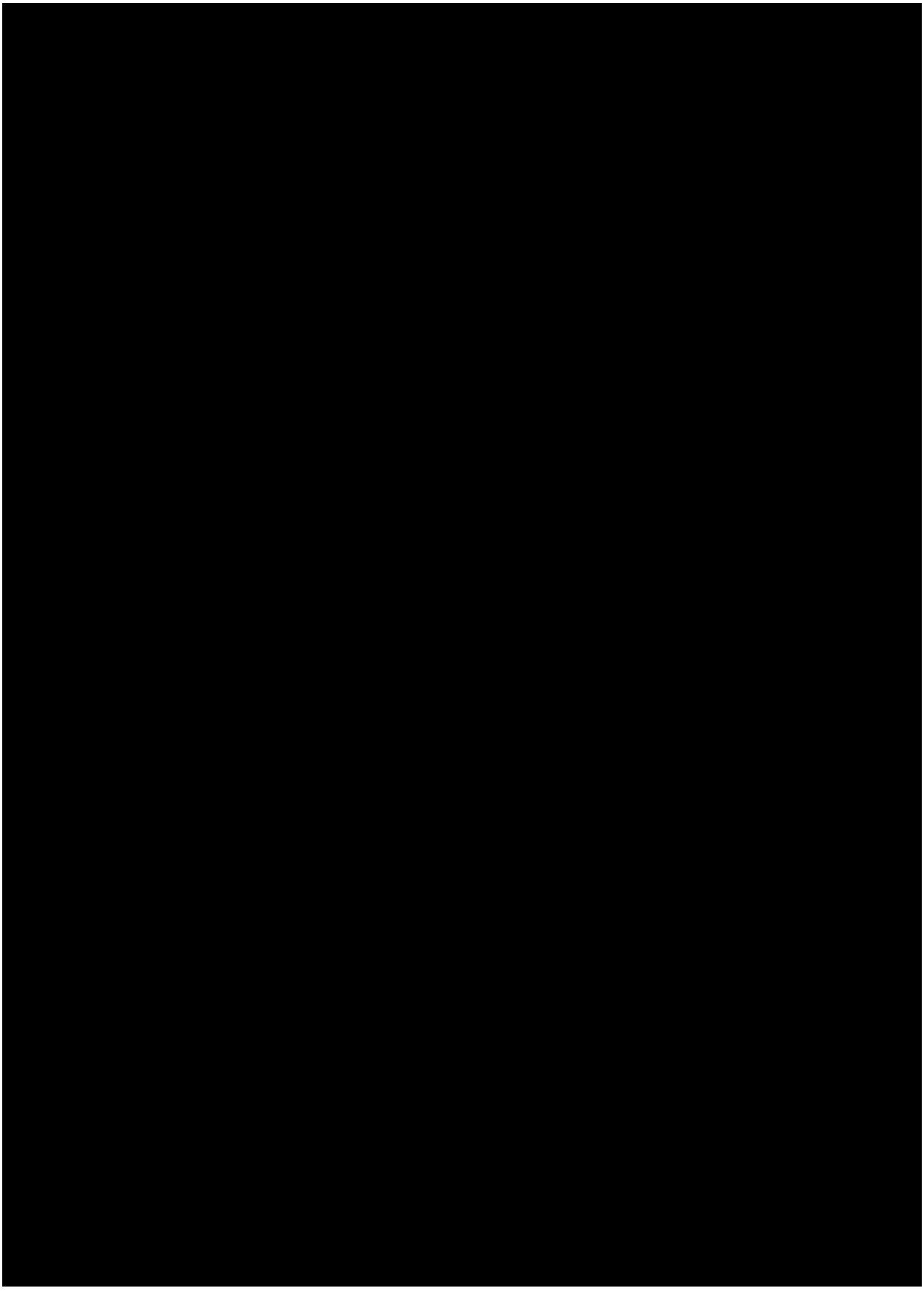
- Supportive analyses on Vineland questionnaires collected during the core phase: to mitigate the impact of many missing data points, analyses of Vineland questionnaires were planned using all available data (including questionnaires retrospectively collected); this analysis related to core phase data was not conducted at the time of the Core database lock because we were not able to retrieve all retrospective questionnaires at the time of the Core database lock; this analysis was described in "CRAD001M2304_M3: Detailed statistical methodology_Final analysis_Addendum 1".
- Long term safety and efficacy analyses described in "CRAD001M2304_M3: Detailed statistical methodology_Final analysis_Addendum 1".
- Additional exploratory long term efficacy analyses described in Section 2.1 of this document.

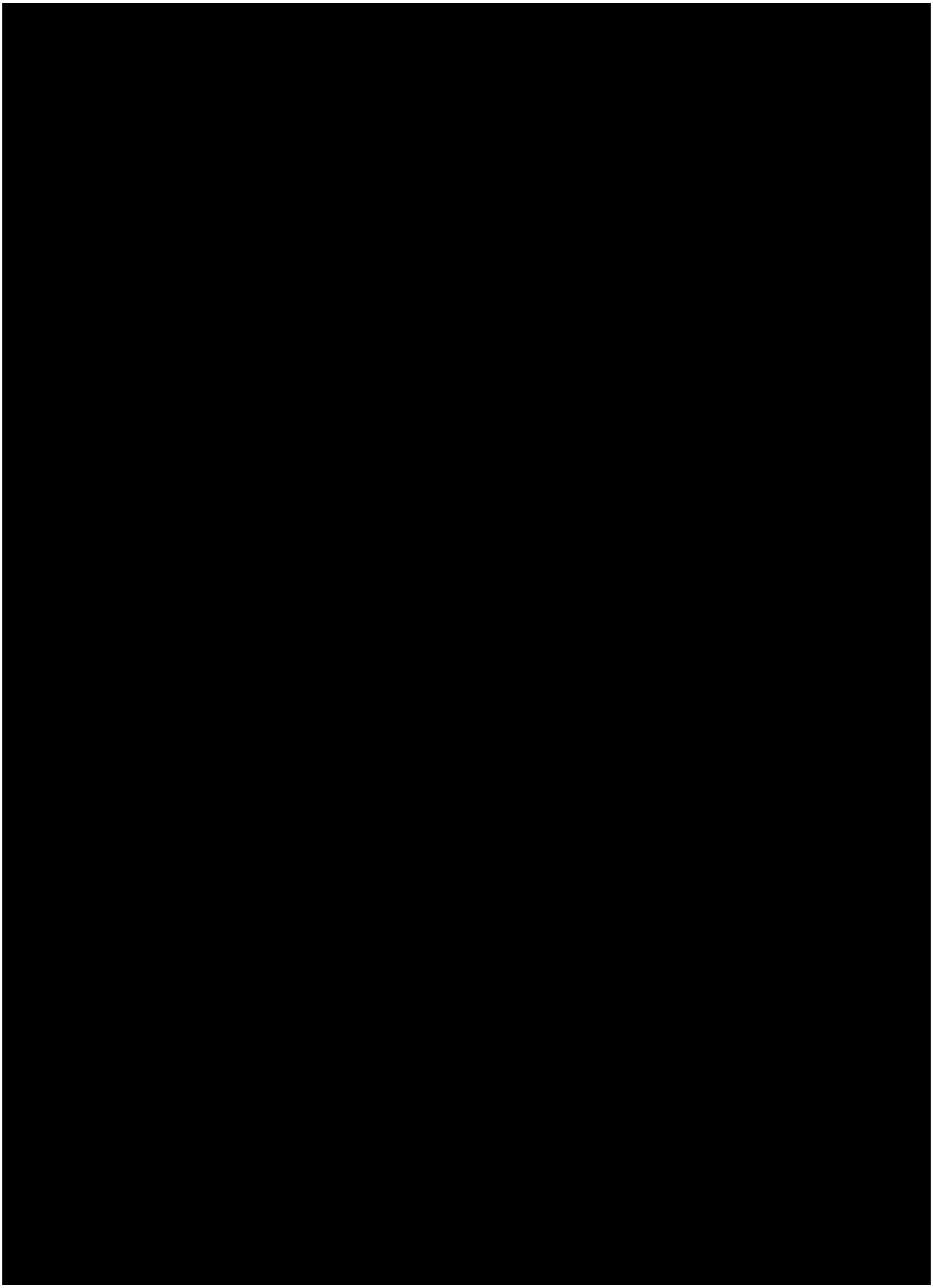
Except for the two first analyses listed above, none of the results focusing on the double-blind Core phase with comparative analyses between everolimus and placebo will be repeated for this CSR.











4 Reference

Not applicable.

NOVARTIS

Clinical Development

Afinitor®/Votubia®/Everolimus

CRAD001M2304 / NCT02962414

A three-arm, randomized, double-blind, placebo-controlled study of the efficacy and safety of two trough-ranges of everolimus as adjunctive therapy in patients with tuberous sclerosis complex (TSC) who have refractory partial-onset seizures

Statistical Analysis Plan (SAP)
to support final Clinical Study Report (CSR)

Author: [redacted]

Document type: SAP Documentation

Document status: Final version

Release date: 11-SEP-2017

Number of pages: 13

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
11-SEP-2017	Prior to DB lock	Creation of final version	N/A - First version	NA

Table of contents

List of abbreviations

AE	Adverse event
AESI	Adverse Event of Special Interest
AED	Antiepileptic drug
CSR	Clinical Study report
EudraCT	European Union Drug Regulatory Authorities Clinical Trial
LTE	Long-Term Evaluation
PEP	Post-Extension Phase
PT	Preferred Term
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCS	Summary of Clinical Safety: preparation document
SOC	System Organ Class
TSC	Tuberous Sclerosis Complex

1 Introduction

Analyses conducted for the Extension CSR (cut-off date of 02SEP2016), focusing on Long-Term Evaluation (LTE) of everolimus using all available data from both the Core and Extension phases of the trial, were described in "CRAD001M2304_M3: Detailed statistical methodology_Final analysis_Addendum 1" (Primary/ Original SAP) and "CRAD001M2304_SAP_Extension CSR_addendum 1" (Extension SAP).

The purpose of this new SAP is to describe the additional analyses planned to further assess the LTE of everolimus in the study population, especially those linked to the Post-Extension Phase (PEP).

The final CSR will include the following analyses:

- Some of the long-term efficacy and safety analyses described in "CRAD001M2304_SAP_Extension CSR_addendum 1" on the LTE Efficacy/Safety populations that were deemed relevant to update with data of patients who were still ongoing in Extension phase at the cut-off date of 02SEP2016. These analyses will use Post-Extension Phase data as well, when applied on the LTE Safety Set.
- Some long-term efficacy and safety analyses on the Long-Term Post-Extension Phase populations (described in the sections 2.3.1, 2.3.2, 2.3.5 and 2.4.7) to assess long-term evaluation of everolimus for patients who continued in PEP specifically. Regarding efficacy, only one analysis will combine data from Extension phase and PEP (section 2.3.2). Other efficacy analyses will either use Core + Extension phases data only or PEP data only. All data from Core, Extension and PEP will be used for the safety part, unless a restriction is specified.
- New additional safety/efficacy analyses on LTE Efficacy/Safety Sets:
    - that are required as per standards for safety clinical trial disclosure registry (2.4.3) or for safety evaluation (2.4.6);
    - that were previously needed for publications purpose (2.3.4) or to address Health Authorities request (2.4.5);
    - or that are slightly modified versions of analyses described in Extension SAP (2.3.3, 2.4.1, 2.4.2) or Summary of Clinical Safety -SCS- (2.4.4); for the purpose of this CSR.

2 Statistical methods

Statistical methods are described in "CRAD001M2304_M3: Detailed statistical methodology_Final analysis_Addendum 1" and "CRAD001M2304_SAP_Extension CSR_addendum 1" except for the additional exploratory analyses described in the section 2.3, 2.4.3, 2.4.4 and 2.4.5.

2.1 Definition of new population analysis sets

Seizure diaries are no longer collected in the Post-Extension Phase, instead, seizure information has been collected as qualitative data through a questionnaire. Seizure diaries were recording, on an ongoing basis up to the end of the Extension, the patient's seizures and their characteristics. From the counts of partial-onset seizures, the primary endpoints were determined. The questionnaire in PEP only gives information on the seizure status of the patient, for instance it only shows yes/no answers to the questions: "was patient seizure-free since last visit?", "have seizures worsened?", "how did they worsen?".

Three new sets of analysis population are therefore defined: "Long-Term Post-Extension Phase Efficacy Set" , "Long-Term Post-Extension Phase Safety Set" and "Long-Term Post-Extension Phase Safety Set- Confirmed PK sample Set".

- LT PEP Efficacy Set includes all patients who received at least one non-zero dose of everolimus during the Post-Extension Phase and with the specific eCRF efficacy page ("Changes in seizures") collected at least once during the Post-Extension Phase.
- LT PEP Safety Set includes all patients who received at least one non-zero dose of everolimus and who have at least one safety evaluation during PEP. LT PEP Safety Set will be used for exposure and concomitant medication related outputs only.
- LT PEP Safety Set-Confirmed PK Sample Set is the same definition as the LTE Safety Set-Confirmed PK Sample Set, except that post-extension phase samples only are used. These two populations are used in Pharmacokinetics/Safety analyses.

Of note, LTE Safety Set, LTE Efficacy Set and LTE Safety Set-Confirmed PK Sample Set definitions have not changed compared to the primary CSR. Only the data used vary according to the analyses (see Table 1 below).

Table 1 Analysis populations and the data they use

Analysis population	Data used
LTE Efficacy Set	Core + Extension (analyses with seizure diaries information)
LT PEP Efficacy Set	Post-Extension (except for analyses described in 2.3.2 & 2.3.5 where Extension data for this population is also included)
LTE Safety set	Core + Extension + Post-Extension (unless it is precised that we consider data up to the end of Extension phase cf 2.4.1)
LT PEP Safety Set	Core + Extension + Post-Extension
LTE Safety Set-Confirmed PK Sample Set	Time windows up to the end of Extension phase
LT PEP Safety Set-Confirmed PK Sample Set	Time windows during PEP

2.2 Background, demographics and disposition additional/modified analyses

2.2.1 Background and demographics

Demographics, Epilepsy background and seizure history, TSC diagnosis, and Seizure history during the baseline phase will be summarized, as they have been defined in Primary SAP, by treatment group and overall but using the LT PEP Efficacy Set.

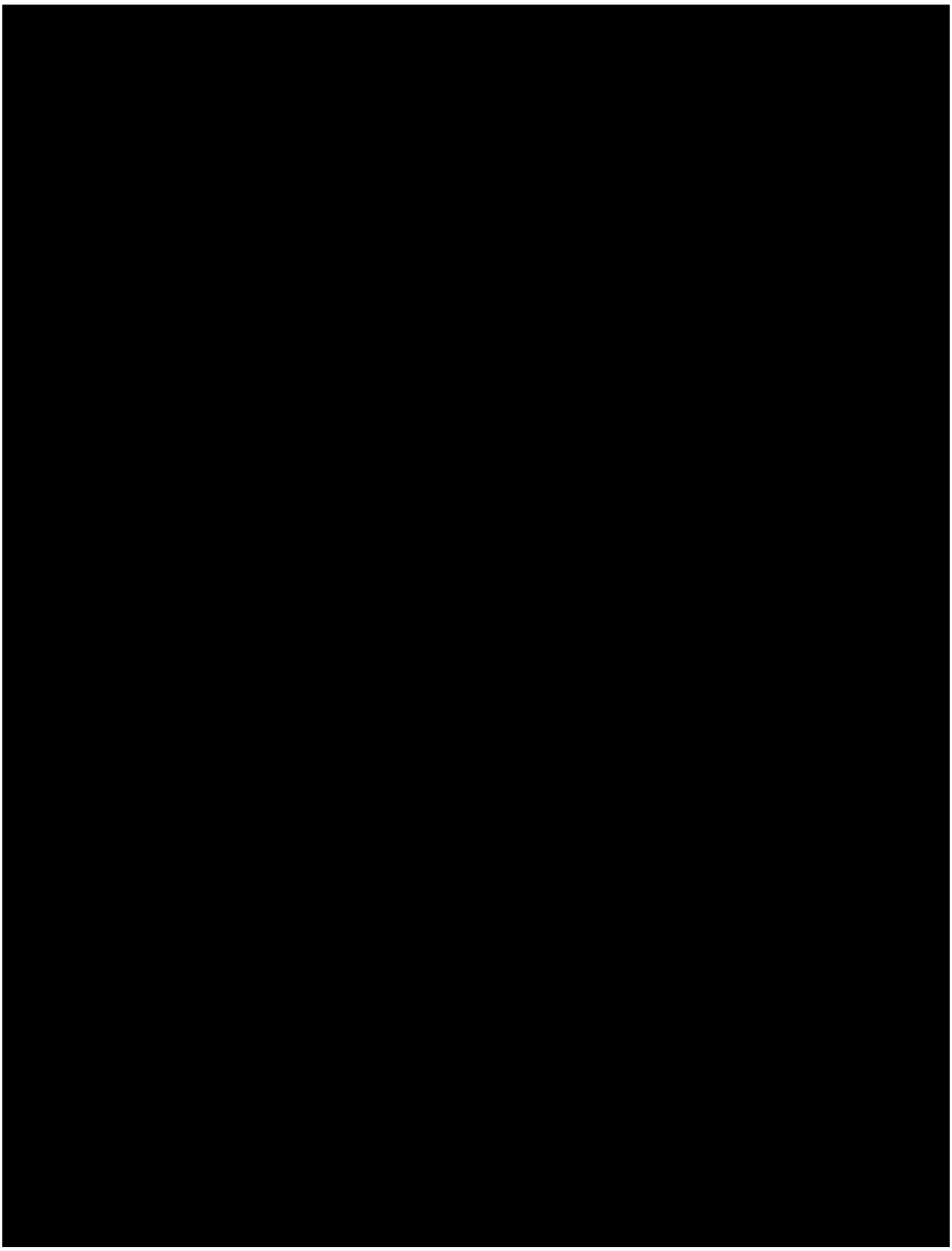
Seizure characteristics at the start of the Post-Extension Phase for patients included in the LT PEP Efficacy set will be presented for each treatment group and overall in order to characterize the seizure profile of patients that have started the PEP.

The start of Post-Extension Phase in this analysis refers to the seizures recorded in the last 12 weeks of the Extension phase.

2.2.2 Patient disposition

In order to better interpret the patient disposition in the different everolimus treatment phases, the analysis suggested in the previous Extension CSR will be modified. The number of patients who have been treated in Core/Extension/Post-Extension phases will be presented separately as well as the number of patients who continued in the corresponding next treatment phase. For patients who did not move into the next treatment phase, the primary reason for end of treatment will be described. The summary of patient disposition will be

performed using LTE Efficacy Set and LT PEP Efficacy Set by treatment arm and over all. The LTE Efficacy Set will be used for the listing.



2.4 Additional/Modified safety analyses

2.4.1 Duration of exposure

In order to link the efficacy results (i.e. percentage seizure reduction from baseline & response rate) with the duration of exposure on one hand, and on the other hand to link the safety results with the duration of exposure, two analyses of duration of exposure will be defined:

- using data up to the end of Extension phase. The duration of exposure will be calculated as described in the original SAP ("CRAD001M2304_M3: Detailed statistical methodology_Final analysis_Addendum 1");
- using data up to the end of Post-Extension Phase. The duration of exposure will be then calculated from the first administration date to the last administration date in the PEP that is defined as the last date when a non-zero dose of everolimus is administered and recorded on the "Dosage administration record – Everolimus – Post-Extension" eCRF.

The duration of exposure will be summarized both categorically and continuously in units of weeks and/or months, by treatment arm and overall on the Long-Term Evaluation Safety Set.

2.4.2 Pharmacokinetics/ Safety analyses

As for the previous analysis, the impact of CYP3A4 inducers and inhibitors on everolimus concentration at trough (Cmin in ng/mL) and dose-normalized everolimus concentration at trough (Cmin in ng/mL per mg and Cmin in ng/mL per mg/m²) as they are defined in section 3.13.1.2 of the Primary SAP, will be assessed descriptively using two different analyses:

- using time windows as defined in the Primary SAP up to the end of Extension phase on the Long-Term Evaluation Safety Set-Confirmed PK Sample Set;
- using time windows calculated only during the Post-Extension-Phase on Long-Term Post-Extension Phase Safety Set-Confirmed PK Sample Set.

Everolimus concentration will be summarized using the two above specifications, by treatment arm. And then separately by treatment arm, age subgroup, use of CYP3A4 inducers and use of CYP3A4 inhibitors. The corresponding boxplots will as well be presented.

2.4.3 On-treatment deaths, serious and non-serious adverse events

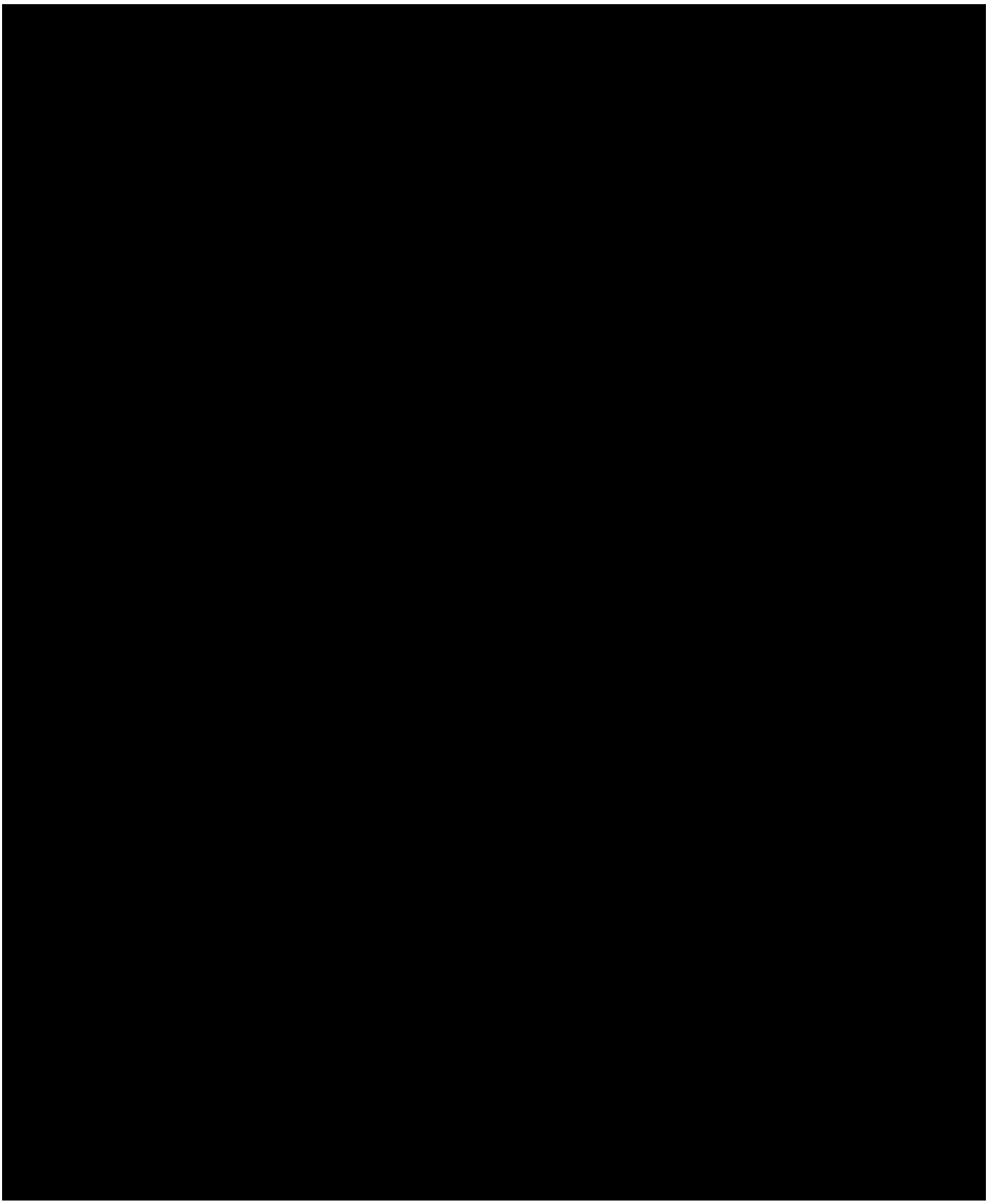
For the legal requirements of ClinicalTrials.gov and EudraCT, two required tables on on-treatment adverse events (AE) which are not serious adverse events (SAE) with an incidence greater than 5% and on on-treatment SAE and SAE suspected to be related to study treatment will be provided by system organ class (SOC) and preferred term (PT) on the LTE Safety Set population. On-treatment assessment is defined as per Primary SAP as an AE started in the following time interval (including the lower and upper limits): [start date of study treatment; date of last study treatment + 30 days].

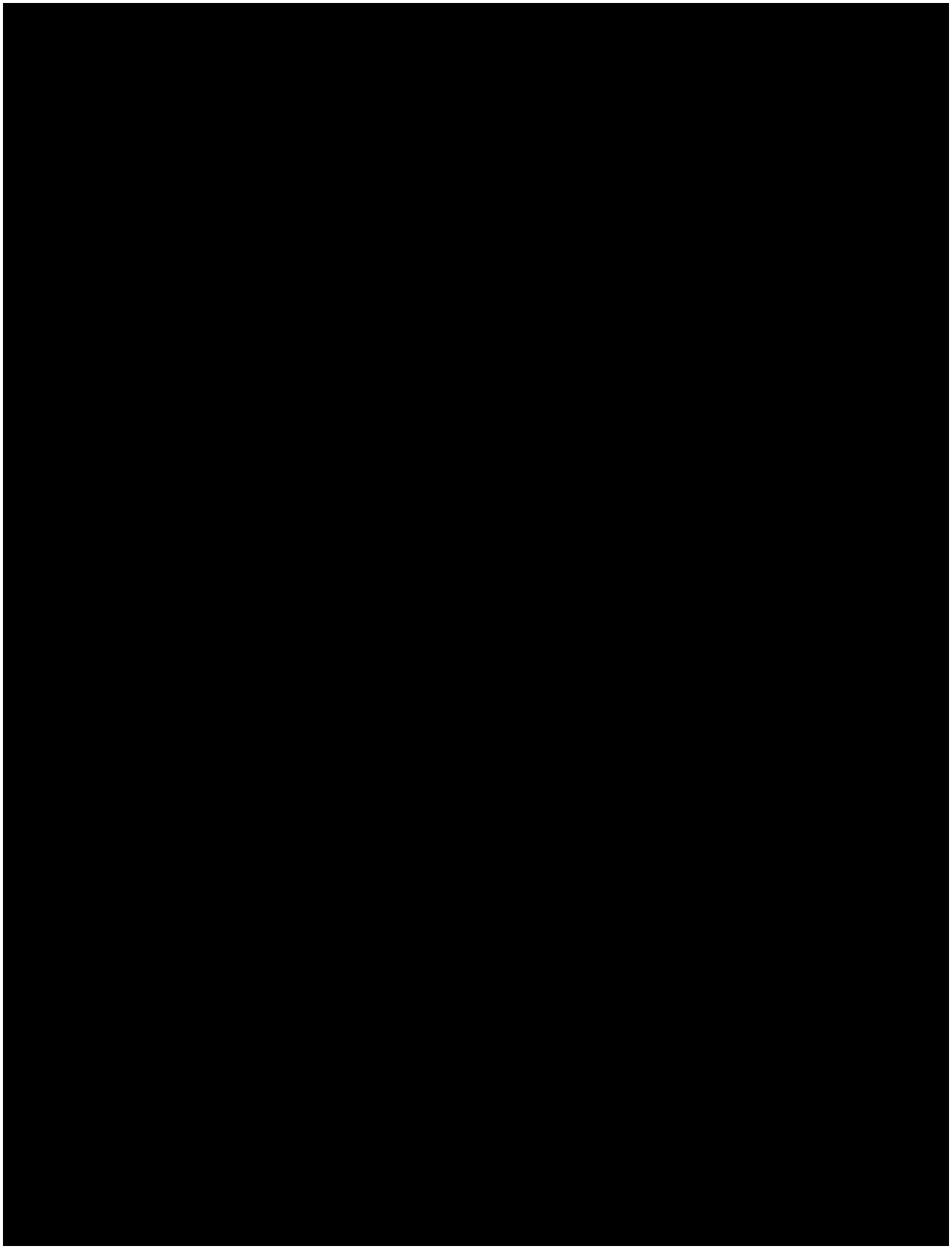
If for a same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

- a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE.
- more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE.

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment /non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.





3 References

Not applicable.