Title: A Phase 2, Randomized, Double-Blind, Placebo-Controlled, Parallel-Arm Study to Evaluate Efficacy, Tolerability, and Pharmacokinetics of Multiple Doses of Oral TAK-831 in Adult Subjects With Friedreich Ataxia

NCT Number: NCT03214588

SAP Approve Date: 01 February 2019

Certain information within this Statistical Analysis Plan has been redacted (ie, specific content is masked irreversibly from view with a black/blue bar) to protect either personally identifiable (PPD) information or company confidential information (CCI).

This may include, but is not limited to, redaction of the following:

- Named persons or organizations associated with the study.
- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.
- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: TAK-831-1501

A Phase 2, Randomized, Double-Blind, Placebo-Controlled, Parallel-Arm Study to Evaluate Efficacy, Tolerability, and Pharmacokinetics of Multiple Doses of Oral TAK-831 in Adult Subjects With Friedreich Ataxia

PHASE 2

Version: Final

Date: 01 February 2019

Prepared by:

PPD

Based on:

Protocol Version: Amendment 1

Protocol Date: 06 October 2017

1.1 Approval Signature

Electronic signature can be found on the last page of this document.

2.0 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF APPENDICES

3.0 LIST OF ABBREVIATIONS

Term	Definition
%CV	coefficient of variation
9-HPT	9-hole peg test
ADaM	Analysis Data Model
ADL	Activities of Daily Living
AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
AUC_{∞}	area under the concentration-time curve from time 0 to infinity
BID	twice daily
BMI	body mass index
CGI	Clinical Global Impression Scale
CGI-I	Clinical Global Impression-Improvement
CGI-S	Clinical Global Impression-Severity
C_{max}	maximum observed concentration
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	electrocardiogram, electrocardiographic
eCRF	electronic case report form
FARS	Friedreich Ataxia Rating Scale
FAS	full analysis set
FRDA	Friedreich ataxia
GGT	gamma-glutamyl transferase
MAV	markedly abnormal value
MedDRA	Medical Dictionary for Regulatory Activities
mFARS	modified Friedreich Ataxia Rating Scale
mFARS-neuro	modified Friedreich Ataxia Rating Scale neurological examination
MMRM	mixed model for repeated measures
PD	pharmacodynamic(s)
PGI-I	Patient Global Impression-Improvement
PGI-S	Patient Global Impression-Severity
PK	pharmacokinetic(s)
PMM	pattern mixture model
PT	preferred term
PTE	pretreatment event
QTc	QT interval with correction
QTcF	QT interval with Fridericia correction method
SAE	serious adverse event
SD	standard deviation
SDTM	Study Data Tabulation Model

Term	Definition
SOC	system organ class
SI	International System of Units
T25FW	Timed 25-Foot Walk
TEAE	treatment-emergent adverse event
ULN	upper limit of normal
WHODrug	World Health Organization Drug Dictionary

4.0 OBJECTIVES

4.1 Primary Objectives

The primary objective of the study is to evaluate the efficacy of TAK-831 versus placebo on upper extremity (arms and hands) motor function and manual dexterity, as measured by the 9-hole peg test (9-HPT).

4.2 Secondary Objectives

- To evaluate the efficacy of TAK-831 versus placebo on Activities of Daily Living (ADL), as measured by the ADL component of the Friedreich Ataxia Rating Scale (FARS), and those items that are related to upper limb function.
- To evaluate the efficacy of TAK-831 versus placebo on neurological function, as measured by the modified FARS neurological examination (mFARS-neuro) total score.
- To evaluate the efficacy of TAK-831 versus placebo on lower extremity motor function, as measured by the Timed 25-Foot Walk (T25FW).
- To evaluate the efficacy of TAK-831 versus placebo on limb ataxia, as measured by the composite of the T25FW and 9-HPT.
- To evaluate the efficacy of TAK-831 versus placebo on visual acuity, as measured by the low-contrast letter acuity (LCLA) testing with Sloan charts.
- To evaluate the efficacy of TAK-831 versus placebo on global improvement, as measured by the Clinical Global Impression-Improvement (CGI-I).
- To evaluate the efficacy of TAK-831 versus placebo on global improvement, as measured by the Patient Global Impression-Improvement (PGI-I).
- To evaluate the efficacy of TAK-831 versus placebo on upper extremity functional improvement, as measured by the CGI-I.
- To evaluate the efficacy of TAK-831 versus placebo on upper extremity functional improvement, as measured by the PGI-I.
- To evaluate the efficacy of TAK-831 versus placebo on global severity, as measured by the Clinical Global Impression-Severity (CGI-S).
- To evaluate the efficacy of TAK-831 versus placebo on global severity, as measured by the Patient Global Impression-Severity (PGI-S).
- To evaluate the efficacy of TAK-831 versus placebo on upper extremity functional severity, as measured by the CGI-S.
- To evaluate the efficacy of TAK-831 versus placebo on upper extremity functional severity, as measured by the PGI-S.
- To evaluate the safety and tolerability of TAK-831.

4.3 Exploratory Objectives

- To evaluate the efficacy of TAK-831 versus placebo on speech, as measured by digital assessments.
- To evaluate the efficacy of TAK-831 versus placebo on motor function such as gait, balance, and upper limb coordination, as measured by digital sensors.
- To evaluate the PK and PD (plasma D-serine and L-serine levels and D-serine:Total serine ratios), and explore the relationships between TAK-831 exposures and D- and L-serine levels or selected measures of efficacy.
- To correlate plasma D-serine and L-serine levels, and D-serine:Total serine ratios with clinical parameters of disease status.
- To quantify subject-perceived meaningful change in upper limb function, as evaluated by the Exit Interview.

CCI

[redacted]

4.4 Study Design

TAK-831-1501 is a phase 2, randomized, double-blind, placebo-controlled, parallel-arm study designed to evaluate the efficacy, safety, PD effects, and PK of 2 dose levels of oral TAK-831 in adult subjects with FRDA.

The study will include a Screening Period (Days -28 to -2), a Training Period (Day -1), Treatment (Days 1 to 84), a Phone Exit Interview (within 7 days from the last dose of study medication of the Double-Blind Treatment Period or within 7 days of notification of termination for the Early Termination Visit), and a Safety Follow-up Phone Call (7 to 17 days after last dose of study drug).

Approximately 65 subjects who meet the study criteria during Screening will be invited to attend training to minimize practice effects on performance-based efficacy assessments on Day -1. On Day 1, subjects who continue to meet inclusion and none of the exclusion criteria will be randomized in a 2:1:2 ratio to TAK-831 high dose, TAK-831 low dose, or placebo, respectively. (If the rate of early withdrawal is substantially greater than predicted, then an increase to the number of randomized subjects may be considered.) Randomization will be stratified by ambulation status (ambulatory vs. nonambulatory).

The study will be conducted in approximately 6 or more specialized academic neurological clinical centers in the United States with expertise in, and access to, patients with FRDA.

Subjects will receive twice daily (BID) study drug dosing or placebo from Day 1 through Day 84. Clinic visits during the treatment period include Day 1, end of Weeks 2, 7, and 12. Clinical assessments, blood collections, and laboratory tests will be conducted at specific time points per the schedule of study procedures (Protocol Appendix A). A semi-structured Exit Interview will be conducted by phone to provide an in-depth, qualitative evaluation of the subject experience and subject-perceived change due to treatment, and a Safety Follow-up Phone Call will be conducted (Table 4.a).

Table 4.a Schematic of Study Design

Screening	Efficacy Assessment Training	Treatment (TAK-831 high dose, TAK-831 low dose, or placebo)		Phone Exit Interview	Safety Follow-up Phone Call (a)
		Randomization, Baseline, and Dosing	Efficacy, PK, PD, and Safety Assessments		
Days -28 to -2	Day -1	Day 1	Day 14 (±3 days), Day 49 (±3 days), Day 84 (±2 days)	≤7 days after Final Visit or termination	7 to 17 days after last dose

(a) The Follow-up Phone Call will occur 7 to 17 days after the last dose of study drug by telephone unless subject report concerning AEs. In these cases, subjects will be brought back into the clinic for re-evaluation at the investigator's discretion.

5.0 ANALYSIS ENDPOINTS

5.1 Primary Endpoint

The primary endpoint for this study is the change from Baseline to Week 12 in the inverse of the time to complete the 9-HPT (9-HPT^{-1}) after treatment with TAK-831, compared with placebo.

5.2 Secondary Endpoints

- The change from Baseline to Weeks 2, 7, and 12 on the ADL component of the FARS, after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on the ADL component upper limb function items (ADL items 3 to 5) of the FARS, after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2 and 7 on the 9-HPT^{-1} after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on the ADL individual items after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on the mFARs-neuro total score after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on the mFARS-neuro subscales and individual items after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on the T25FW after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on the 9-HPT and T25FW composite score after treatment with TAK-831, compared with placebo.
- Percentage of subjects whose 9-HPT completion time is reduced by at least 15% and at least 20% from Baseline.
- The change from Baseline to Weeks 2, 7, and 12 on LCLA test score after treatment, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on CGI-I (global change) after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on PGI-I (global change) after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on CGI-I (upper extremity functional change) after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on PGI- I (upper extremity functional change) after treatment with TAK-831, compared with placebo.

- The change from Baseline to Weeks 2, 7, and 12 on CGI-S (global severity) after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on PGI-S (global severity) after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on CGI-S (upper extremity functional severity) after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on PGI-S (upper extremity functional severity) after treatment with TAK-831, compared with placebo.

5.3 Safety Endpoints

- AEs, clinical laboratory assessments, ECG measures, changes in physical examinations and vital signs after treatment with TAK-831, compared with placebo.
- Assessment of suicidal ideation and behavior as measured by the C-SSRS at Screening, Baseline, and all other timepoints after treatment with TAK-831, compared with placebo.

5.4 Exploratory Endpoints

- The change from Baseline to Weeks 2, 7, and 12 on the digital speech assessment after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on the digital motor sensors measurements such as gait, stability and upper limb coordination after treatment with TAK-831, compared with placebo.
- The change from Baseline to Weeks 2, 7, and 12 on TAK-831 plasma D-serine and L-serine levels and D-serine:Total serine ratios after treatment with TAK-831, compared with placebo.
- Percent of subjects reporting a meaningful change in upper limb function from Baseline to Week 12 after treatment with TAK-831, compared with placebo, as evaluated by the Exit Interview.
- Correlation of change in 9-HPT with that of patient reported improvement of change in upper limb function, such as ADL items 3 to 5 and PGI.
- TAK-831 plasma concentration.

6.0 DETERMINATION OF SAMPLE SIZE

Approximately 65 subjects will be randomized in a 2:1:2 ratio for the placebo, low, and high dose respectively. It is assumed that approximately 60 subjects (92%) will complete 12 weeks and be evaluable on the primary endpoint. If the rate of early withdrawal is substantially greater than predicted, then an increase to the number of randomized subjects may be considered.

Assuming an effect size of 0.6 for each dose, there will be 77% power for at least 1 dose to be deemed superior to placebo and 54% power for both doses (overall alpha=0.10, 1-sided). Validated software (EAST 6.3, Cytel Inc.) was used to determine the power. The effect size was chosen based on the premise that a treatment effect equivalent to preventing or reversing 2 years of decline due to FDRA would be clinically meaningful. Using published data (Patel et al., 2016 [2]), the target effect size is approximately 0.6 for the 9-HPT.

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 General Principles

Baseline values are defined as the last observed value before the first dose of study medication.

All confidence intervals, statistical tests, and resulting P-values will be reported as 2-sided and will be assessed at α=0.05 significance level unless otherwise specified. P-values will be rounded to 3 decimal places prior to assessment of statistical significance.

Unless otherwise stated, means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. Confidence intervals about a parameter estimate will be presented using the same number of decimal places as the parameter estimate.

Continuous data will be summarized using the following descriptive statistics: number of subjects (N), mean, standard deviation (SD), median, minimum, and maximum, where appropriate. Where indicated, the coefficient of variation (%CV) and geometric mean will also be included in the summary of continuous data. Categorical data will be summarized using the number and percent of subjects for each category, where appropriate.

Unless otherwise stated, Baseline is defined as the last observed value before the first dose of study medication.

As applicable, summaries will be presented by treatment arm, TAK-831 overall and overall. Where appropriate, variables will be summarized descriptively by study visit.

All data analyses and figures will be generated using SAS System® Version 9.4 or higher.

7.1.1 Study Definitions

7.1.1.1 Definition of Study Days

Study Day 1 is defined as the date on which a subject is administered their first dose of the medication. Other study days are defined relative to the Study Day 1 with Day 1 being Study Day 1 and Day -1 being the day prior to Study Day 1.

7.1.1.2 Definition of Study Visit Windows

A windowing convention will be used to determine the analysis value for a given study visit that applies to observed data. For each visit, a window will be defined such that the lower and upper bounds of each window is generally the midpoint between 2 consecutive study visits.

More than 1 result for a parameter may be obtained in a visit window. In such an event, the result with the date closest to the scheduled visit day will be used. In the event of 2 observations equidistant to the scheduled visit day, the later of the observations will be used.

7.1.2 Conventions for Missing Adverse Event Dates

Incomplete adverse event (AE) start dates will be imputed to determine the relationship between the start date and the informed consent date, as well as the start date and the first dose date of the double-blind study medication. Incomplete AE dates will be presented as they are in the listings.

The following methods will be used to impute incomplete start dates of AEs:

- If only the month and year of the start date are available and the month and year are different than the month and year of the first dose of double-blind study medication or the stop date is prior to the first dose of study medication, then the first day of the month will be used for the start date. If only the month and year of the start date are available and the month and year are the same as the month and year of the first dose of double-blind study medication and the stop date is not prior to the date of first dose, then the date of first dose will be used for the start date.
- If only the year of the start date is available and the year is different than the year of the first dose of double-blind study medication or the stop date is prior to the first dose of study medication, then January 1st will be used for start date. If only the year of the start date is available and the year is the same as the year of the first dose of double-blind study medication and the stop date is not prior to the date of first dose, then the date of first dose will be used for start date.

7.1.3 Conventions for Missing Concomitant Medication Dates

If start date and stop date are missing, medication will be assumed to occur both prior and concomitantly.

7.1.4 Grouping of Sites

Before unblinding the data, study sites that are considered small (< 5 subjects for analysis) will be pooled with geographically similar sites to minimize artifacts in the statistical analyses from imbalances in subject counts within the sites.

The pooling will start first within each state; if it is still small, other states within the same geographical region will also be considered. If possible, sites will only be pooled with sites from within the same general US geographic region.

The pooling of the sites will be reviewed before the final analysis.

7.2 Analysis Sets

7.2.1 Full Analysis Set

The full analysis set (FAS) will include all subjects who were randomized and received at least 1 dose of the study drug for the treatment period. In FAS efficacy summaries, subjects will be analyzed by the treatment to which they were randomized.

7.2.2 Safety Analysis Set

The Safety Analysis Set will include all randomized subjects who receive at least 1 dose of double-blind study medication. In safety summaries, subjects will be analyzed according to the treatment they received.

7.2.3 PD Analysis Set

The PD analysis set will consist of subjects who receive at least 1 dose of double-blind study medication and who have a Baseline D-serine and L-serine plasma concentration measurement and at least corresponding 1 postdose plasma concentration measurement.

7.2.4 PK Analysis Set

The PK analysis set will consist of subjects who receive at least 1 dose of double-blind study medication and who have any available plasma concentration data.

7.3 Disposition of Subjects

Disposition of all screened subjects (denominator) will be tabulated (count and percent); there will be no inferential analysis of subject disposition data.

Summaries will be presented by treatment arm, TAK-831 overall and overall.

Disposition of all randomized subjects will be tabulated:

- All subjects received at least one dose of study drug (denominator).
- Subjects who completed the study.
- Subjects who prematurely discontinued study.

Primary reasons for discontinuation of study and study treatment as entered on the electronic case report form (eCRF), will be tabulated. Reasons for discontinuation include death, adverse event, significant protocol deviation, lost to follow-up, withdrawal by subject, study termination by sponsor, pregnancy, severe gastrointestinal intolerance and other. The date of first dose, date of last dose, duration of treatment and the reason for premature discontinuation of study drug/study visit will be presented for each subject in listings.

Disposition of screen failure subjects will be summarized descriptively. Primary reasons for failure will be summarized and will be presented in a data listing.

Significant protocol deviation will be summarized by type.

7.4 Demographic and Other Baseline Characteristics

Demographics and baseline characteristics including age at screening, age at disease onset, gender, race, and BMI will be listed and summarized by each treatment group and overall based on the FAS.

Baseline values for efficacy assessments will also be summarized by each treatment group and overall based on the FAS.

For continuous variables, comparability of treatment groups will be assessed using an analysis of variance with treatment and pooled site as factors. For discrete variables, comparability will be assessed using the Cochran-Mantel-Haenszel general association test, stratified by pooled site. The statistical test for race will only be conducted if < 90% subjects report the same race.

7.5 Medical History and Concurrent Medical Conditions

Medical history and concurrent medical conditions will be coded using the Medical Dictionary for Drug Regulatory Activities (MedDRA, version 21 or higher) coding system.

Medical history to be obtained will include determining whether the subject has any significant conditions or diseases relevant to FRDA (including age at disease onset) that resolved at or prior to signing of informed consent. Ongoing conditions are considered concurrent medical conditions.

All medical history and concurrent medical condition data will be listed by site (study center) and subject number. The listing will contain subject identifier, treatment, system organ class (SOC), preferred term (PT), start and end dates (if not ongoing), and the verbatim term.

Medical history and concurrent medical conditions will be listed and summarized by system organ class and preferred term for each treatment group and overall based on the FAS.

7.6 Medication History and Concomitant Medications

Medication history information to be obtained includes any medication relevant to eligibility criteria. Concomitant medication is any drug given in addition to the study drug.

All medication history and concomitant medications will be listed by site (study center) and subject number. The listings will contain subject identifier, treatment, World Health Organization Drug Dictionary (WHODrug) preferred medication name, dose, unit, frequency, route, start date, stop date, whether the medication was ongoing, and reason for use. Medication history and concomitant medications will be summarized for each treatment group and overall based on the FAS. No inferential statistics will be presented.

Medication history and concomitant medications will be coded using the WHODrug Version 01March2017 or higher.

7.7 Study Drug Exposure and Compliance

Duration of exposure is defined as [date of the last dose of study medication - date of first dose of study medication + 1]. Compliance will be defined as {(number of tablets dispensed – number of tablets returned)/(duration of exposure in days x 2 x 3)} x 100%. Duration of exposure and compliance will be summarized for each treatment group and overall based on the FAS. The duration of exposure will be summarized as a continuous variable.

7.8 Efficacy Analysis

Two 9-HPT assessments are conducted at each assessment date, with each assessment consisting of two trials for each hand. The 9-HPT^{-1} endpoint will be computed as follows for each assessment, using the formula from the Multiple Sclerosis Functional Composite Administration and Scoring Manual (Fischer et al, 2001 [1]):

1) For the dominant hand, compute the mean of the two trials, and take the reciprocal
2) Do the same for the non-dominant hand
3) Take the average of the two reciprocals

For descriptive summaries of the 9-HPT, the overall value for each subject will be defined as the inverse of the 9-HPT^{-1} (as computed above). Except where otherwise noted, the analyses described herein will be on the data from the first assessment.

Descriptive statistics (mean and SD) for change in the 9-HPT^{-1} will be displayed with at least 5 digits after the decimal point.

7.8.1 Primary Efficacy Endpoint

The primary endpoint for this study is the change from Baseline to Week 12 in the inverse of the time to complete the 9-HPT (9-HPT^{-1}) after treatment with TAK-831, compared with placebo.

The primary analysis of change from Baseline in 9-HPT^{-1} will be based on the FAS. Comparisons between TAK-831 and placebo will be made over all assessed time points using estimates from a mixed model for repeated measures (MMRM) with Baseline 9-HPT^{-1} as a covariate; pooled site, visit, treatment, and ambulation status (randomization factor) as fixed factors; and treatment-by-visit and Baseline 9-HPT^{-1}-by-visit interactions. Based on a Missing at Random Assumption, this analysis will be performed using observed case data only. The effect at each time point for each treatment is allowed to vary freely and an unstructured covariance matrix is assumed. If the model does not converge, other covariance structures will be considered in the following order, from most generous to most parsimonious: Toeplitz, compound symmetry, and Toeplitz with two bands.

Since 2 doses of TAK-831 are compared to placebo, multiplicity will be controlled at the Holm's method. The dose with the smaller p-value will be tested at the one-sided 0.05 level, then dose with the larger p-value will be tested at the one-sided 0.10 level only if the smaller p-value is significant. This procedure will control the overall type I error rate at the one-sided 0.10 level. The one-sided alternative hypothesis is that the mean time to complete the 9-HPT is shorter on the TAK-831 arm compared to placebo.

Subgroup analyses will be conducted according to ambulation status, Baseline D-serine concentration (< median of randomized subjects and ≥ median), sex, categorical age at Screening (<35 and ≥35 years of age), and baseline mFARS-neuro total score (< median of randomized subjects and ≥ median). In addition, a subgroup analysis by race (white vs. non-white) will be conducted if at least 20% of randomized subjects are non-white.

As a sensitivity analysis, a method for missing value imputation based on pattern mixture models (PMM) will be implemented using standard SAS STAT procedures. It is assumed that after discontinuation from the study, subjects receiving a TAK-831 dose will exhibit the same future evolution of the disease as subjects on the placebo treatment. The missing 9-HPT^{-1} score in the TAK-831 dose arm will be imputed using placebo-based pattern mixture models. This method uses sequential regression and multiple imputation methodology to impute missing values for visits after a subject's discontinuation from the study in the placebo arm. Then based on the available data from placebo subjects, the missing values from active treatment arm will be imputed using pattern mixture models until missing values at all visits are imputed.

The primary endpoint will then be analyzed using ANCOVA, with pooled center, stratification factor, and treatment as fixed factors, and Baseline 9-HPT^{-1} as a covariate, based on the multiply-imputed data.

In addition, the primary analysis will be conducted with the average of Screening and Baseline 9-HPT^{-1} included in the model instead of Baseline 9-HPT^{-1}.

The primary analysis will also be conducted using the second set of 9-HPT assessments at each timepoint. Both sets of 9-HPT assessments will be summarized descriptively at all timepoints.

The 9-HPT and 9-HPT^{-1} will be plotted longitudinally for each subject. Up to 10 subjects from the same treatment group can be displayed on the same plot.

7.8.2 Secondary Efficacy Endpoints

Analysis of secondary efficacy endpoints will be based on the FAS. For these endpoints, which are continuous change from Baseline endpoints, comparisons between TAK-831 and placebo will be based on similar methodology to that described for the primary efficacy analysis. For binary or ordinal endpoints, Cochran-Mantel-Haenszel tests will be used with ambulation status at randomization as a stratification factor. For binary or ordinal endpoints, missing data will be imputed as no change from Baseline. For each secondary endpoint, the 1-sided alternative hypothesis is that TAK-831 is superior to placebo in the clinically favorable direction, for example, a lower mean score on the FARS ADL scale.

The time-weighted average of the change from baseline in 9-HPT^{-1} at Weeks 2, 7, and 12 will be summarized for each treatment group. Specifically, the change from baseline will be weighted according to the number of days since the first dose (at Week 2) or the previous post-baseline observation (at Weeks 7 and 12).

The T25FW assessment will be consist of up to 2 trials at each assessment. If both are conducted, the average of the 2 trials will be used in the analysis.

The 9-HPT^{-1} and T25FW composite score will be created using the methods in Fischer et al (2001). In brief, raw scores from each test will be converted to test-specific Z scores by subtracting the cohort mean from the raw score, and then dividing by the cohort SD. The cohort mean and SD will be computed from the Baseline value, pooling all treatment arms. The composite Z scores will be created by subtracting Z-score for T25FW from the Z-score for 9-HPT^{-1}.

The T25FW and the inverse of the T25FW will be plotted longitudinally for each subject. Up to 10 subjects from the same treatment group can be displayed on the same plot. Plotting symbols will be used to indicate what type of assistive device, if any, a subject used at each assessment.

Multiplicity will not be controlled for the secondary endpoints.

7.8.3 Exploratory Endpoints

The analysis of the digital speech and motor endpoints will be described in a separate analysis plan.

7.9 Pharmacokinetic/Pharmacodynamic Analysis

7.9.1 Pharmacokinetic Analysis

Plasma concentrations of TAK-831 will be listed for each subject and summarized by each time point for each dose of the study.

Individual concentration-time data will be pooled to describe the population PK of TAK-831. As data permit, a nonlinear mixed effects modeling approach (NONMEM software) will be used to assess TAK-831 exposure. PK information generated in this study will be further utilized in subsequent population PK-PD analyses. The relationships between TAK-831 plasma concentrations and drug response (D- or L-serine plasma levels and/or selected measures of efficacy) will be explored. As appropriate, historical data may be used in this analysis to increase the robustness of the model and precision of estimated parameters. Details of the modeling approach will be provided in a separate analysis plan, and the results of these analyses may be reported separately.

7.9.2 Pharmacodynamic Analysis

For each regimen, the concentrations of D-serine, L-serine, and the ratio of D-serine to total serine with change will be summarized at each time point of each dose level using descriptive statistics. In addition, a mixed effects regression model will be fitted to the change from Baseline in these concentrations. Comparisons between TAK-831 and placebo will be made over all assessed time points using estimates from a MMRM with serine (D-, L-, ratio) as a covariate; visit and treatment as fixed factors; and treatment-by-visit and Baseline-by-visit interactions. Pairwise comparisons between each pair of test regimens (high dose, low dose and placebo) will be made and the CIs for the difference in the LS means will be constructed for Days 14, 49, and 84. In addition, the percent change from Baseline will be summarized at Day 84.

See also Section 7.9.1.

7.10 Other Outcomes

The CGI-I, PGI-I, CGI-S, and PGI-S endpoints, including both global and upper extremity function endpoints, will be summarized using descriptive statistics.

The statistical details of the Responder Analysis (Protocol Section 13.1.7) will be included in a separate analysis plan.

7.11 Safety Analysis

Analysis of safety variables will be based on the safety analysis set.

7.11.1 Adverse Events

Verbatim terms will be coded by SOC and PT using MedDRA (version 20.0 or later).

A PTE is defined as any untoward medical occurrence in a clinical investigation subject who has signed informed consent to participate in a study but prior to administration of any study drug; it does not necessarily have to have a causal relationship with study participation.

Treatment-emergent adverse events (TEAEs) will be defined as any sign, symptom, syndrome, or new illness, regardless of relationship to study drug, that occur after the first dose of study drug and up to 3 days (onset date – last date of dose + 1 ≤ 3) after the last dose of study drug or early termination. In the protocol they are referred to simply as ***adverse events***, but the term ***TEAE*** is used in this document and the resulting outputs for consistency with other TAK-831 studies.

TEAEs are recorded in the eCRF as being related or not related to study drug and study procedure. TEAEs that are recorded as related to study drug and/or study procedure will be summarized separately. TEAEs will also be presented by intensity/severity (mild, moderate, and severe). Serious TEAEs, TEAEs leading to study drug discontinuation, and TEAEs leading to death will also be summarized using SOC and PT.

When calculating the frequency and percentage of subjects who reported TEAEs, a subject will be counted only once for each SOC or PT when multiple TEAEs are coded to the same SOC or PT. For the intensity or relatedness summaries, if a subject reports multiple TEAEs coded to the same SOC or PT, the TEAE with maximum intensity or strongest relationship will be included in the summary.

AEs with missing intensity will be listed as such in the AE listings, however, will be summarized as severe in summary tables. If the relationship of an event is missing, the relationship for the event will be considered to have been related.

In general, AEs will be tabulated at the following levels: overall summary (subjects with at least 1 AE in any dose or regimen), the MedDRA SOC, and the MedDRA PT. The tables will include the number and percentage (N[%]) of subjects. Summary tables that will be generated will include, but may not be limited to:

- Overview of Treatment-Emergent Adverse Events
- Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
- Treatment-Emergent Adverse Events by Preferred Term

- Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
- Most Frequent (>5% or N>2) Non-Serious Adverse Events by System Organ Class and Preferred Term
- Relationship of Treatment-Emergent Adverse Events to Study Drug by System Organ Class and Preferred Term
- Drug-Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
- Drug-Related Treatment-Emergent Adverse Events by Preferred Term
- Intensity of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
- Treatment-Emergent Adverse Events Leading to Study Drug Discontinuation by System Organ Class and Preferred Term
- Pretreatment Adverse Events by System Organ Class and Preferred Term

In addition, subject mappings for the TEAEs by SOC and PT will be generated.

Data listings will be provided for PTEs, TEAEs, TEAEs leading to study drug discontinuation, liver function abnormalities, SAEs, and AEs that resulted in death. AEs happened after 3 days post the last dose of the study drug will be listed as well.

Summaries will be presented by treatment arm, TAK-831 overall and overall.

7.11.2 Clinical Laboratory Evaluations

Clinical laboratory tests will be assessed using the Safety Set and will be evaluated and presented using International System of Units (SI) units unless otherwise stated. Refer to Protocol Section 9.1.10 as well as the schedule of the events for a list of all clinical laboratory tests.

All laboratory test parameters will be displayed in individual subject data listings. For test results not in SI units, the conversion to SI units will be done in the derived SDTM and ADaM datasets using the known conversion factors. If necessary, SI units from the central laboratory may be converted to Takeda's preferred SI units in the derived SDTM and ADaM datasets. All summaries and analyses will be based on the values using these preferred SI units.

Only observations within 3 days of the last dose of study drug will be included in the tables. No inferential statistics will be presented unless otherwise stated.

Descriptive statistics (N, mean, SD, median, minimum, and maximum) for the observed and change from baseline values will be presented. Study baseline will be used for change from baseline. Note that "character" urinalysis tests will only be listed.

Laboratory MAVs, identified by the criteria defined in Appendix A, will be tabulated. If a subject has a MAV for a particular laboratory test, all visits for that subject for that parameter will be listed. The number and percentage of subjects with at least 1 postdose markedly abnormal laboratory test result will be summarized. The mapping of the subjects who meet the MAV

criteria will be listed as a table. All observations, including ones at unscheduled visits, will be included in the MAV evaluation and summaries.

Listings of all clinical safety laboratory data will be provided in Appendix 16.2. Laboratory data outside of the normal reference range will be indicated in the listings. In addition, MAVs will be flagged. The listing will include site number, subject identifier, age, gender, treatment group, study visit, and sample collection date.

7.11.3 Vital Signs

Descriptive statistics (N, mean, SD, median, minimum, and maximum) will be used to summarize vital sign parameters at baseline, each post-baseline visit, and change from baseline to each post-baseline visit. Only observations within 3 days of the study drug will be included in the tables.

Vital sign MAVs, identified by the criteria defined in Appendix B, will be tabulated. If a subject has a MAV for a particular vital signs parameter, all visits for that subject for that parameter will be listed. The number and percentage of subjects with at least 1 postdose markedly abnormal vital signs measurement will be summarized. The mapping of the subjects who meet the MAV criteria will be listed as a table. All observations, including ones at unscheduled visits, will be included in the MAV evaluation and summaries.

Listings of all vital signs data will be provided in Appendix 16.2. Vital sign MAVs will be flagged in the listings. The listing will include site number, subject identifier, age, gender, treatment group, study visit, and sample collection date.

7.11.4 12-Lead ECGs

Descriptive statistics (N, mean, SD, median, minimum, and maximum) of ECG parameters, including heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc interval (Fredericia's and Bazett's corrections), will be presented for baseline, each post-baseline visit, and change from baseline to each post-baseline visit. Only the scheduled measurements will be included in the summary. Only observations within 3 days of the study drug will be included in the tables. No inferential statistics will be presented.

ECG MAVs, identified by the criteria defined in Appendix C, will be tabulated. If a subject has a MAV for a particular 12-lead ECG parameter, all visits for that subject for that parameter will be listed. The number and percentage of subjects with at least 1 postdose markedly abnormal 12-lead ECG measurement will be summarized. The mapping of the subjects who meet the MAV criteria will be listed as a table. All observations, including ones at unscheduled visits, will be included in the MAV evaluation and summaries.

Overall ECG interpretation category (normal, abnormal not clinically significant, abnormal clinically significant) is collected by eCRF at baseline and at each scheduled post-baseline visit. Shifts in ECG interpretation will be presented as cross-tabulations (baseline versus each post-baseline visit) of numbers of subjects with normal, abnormal not clinically significant, and

abnormal clinically significant interpretations with missing, if applicable, and total categories by regimen.

Listings of all 12-lead ECG data will be provided in Appendix 16.2. MAVs will be flagged in the listings. The listing will include site number, subject identifier, age, gender, treatment group, study visit, and sample collection date.

The Frederica-corrected QT interval (QTcF) will be derived using the formula in Section 9.1.13 of the protocol. The Frederica-corrected QT interval (QTcB) will be derived using the following formula:

$$QTcB = \frac{QT}{\sqrt{RR}}.$$

7.11.5 Other Observations Related to Safety

Suicidality assessments (C-SSRS) will also be summarized for each treatment group. Subjects with elevated liver function test values will also be summarized.

7.12 Interim Analysis

No interim analysis of efficacy is planned.

The sponsor established an internal DMC independent of the study and project team to periodically review unblinded ***safety*** data during conduct of this study to complement the routine safety monitoring approach for compounds at this stage of development. The functions and procedures of the committee were outlined in a DMC Charter.

7.13 Changes in the Statistical Analysis Plan

Although change in CGI-I and PGI-I were included as endpoints in the protocol, CGI-I and PGI-I were not collected at baseline since these endpoints are not meaningful prior to treatment. Therefor the CGI-I or PGI-I value at the given timepoint rather than the change will be analyzed.

8.0 REFERENCES

1. Fischer JS, Jak AJ, Kniker JE, Rudick RA, Cutter G. Multiple Sclerosis Functional Composite administration and scoring manual, Revised. October 2001. National Multiple Sclerosis Society.
2. Patel M, Isaacs CJ, Seyer L, Brigatti K, Gelbard S, Strawser C, et al. Progression of Friedreich ataxia: quantitative characterization over 5 years. Ann Clin Transl Neurol 2016;3(9):684-94.

Appendix A Criteria for Identification of Markedly Abnormal Laboratory Values

Hematology—Criteria for Markedly Abnormal Values

Parameter	Unit	Low Abnormal	High Abnormal
Hemoglobin	Both	< 0.8 × LLN	> 1.2 × ULN
Hematocrit	Both	< 0.8 × LLN	> 1.2 × ULN
RBC count	Both	< 0.8 × LLN	> 1.2 × ULN
WBC count	Both	<0.5 x LLN	>1.5 x ULN
Platelet count	Conventional	<75 x 10^3/μL	>600 x 10^3/μL
	SI	<75 x 10^9/L	>600 x 10^9/L

LLN=lower limit of normal, RBC=red blood cell, ULN=upper limit of normal, WBC=white blood cell.

Serum Chemistry—Criteria for Markedly Abnormal Values

Parameter	Unit	Low Abnormal	High Abnormal
ALT	Both	--	>3x ULN
AST	Both	--	>3x ULN
GGT	Both	--	>3x ULN
Alkaline phosphatase	Both	--	>3x ULN
Total bilirubin	Conventional	--	>2.0 mg/dL
	SI	--	>34.2 μmol/L
Albumin	Conventional	<2.5 g/dL	--
	SI	<25 g/L	--
Total protein	Both	<0.8x LLN	>1.2x ULN
Creatinine	Conventional	--	>2.0 mg/dL
	SI		>177 μmol/L
Blood urea nitrogen	Conventional		>30 mg/dL
	SI		>10.7 mmol/L
Sodium	Conventional	<130 mEq/L	>150 mEq/L
	SI	<130 mmol/L	>150 mmol/L
Potassium	Conventional	<3.0 mEq/L	>6.0 mEq/L
	SI	<3.0 mmol/L	>6.0 mmol/L
CPK	Both	--	>5x ULN
Glucose	Conventional	< 50 mg/dL	>350 mg/dL
	SI	< 2.8 mmol/L	>19.4 mmol/L

ALT=alanine aminotransferase, AST=aspartate aminotransferase, CPK=creatine phosphokinase, GGT=γ-glutamyl transferase, LLN=lower limit of normal, ULN=upper limit of normal.

Appendix B Criteria for Abnormal Vital Signs

Parameter	Unit	Lower Criteria	Upper Criteria
Pulse	Bpm	<50	>120
Systolic blood pressure	mm Hg	<85	>180
Diastolic blood pressure	mm Hg	<50	>110
Body temperature	^{o}C	< 35.6	>37.7

Appendix C Criteria for Out-of-Range Values for the 12-Lead ECG Parameters

Parameter	Lower Criteria	Upper Criteria
Heart rate	<50 beats per minute	>120 beats per minute
PR	≤80 milliseconds	≥200 milliseconds
QT Interval	≤300 milliseconds	≥460 milliseconds
QTcB Interval	≤300 milliseconds	≥500 milliseconds OR ≥30 milliseconds change from baseline and ≥450 milliseconds
QTcF Interval	≤300 milliseconds	≥500 milliseconds OR ≥30 milliseconds change from baseline and ≥450 milliseconds
QRS	≤80 milliseconds	≥180 milliseconds

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'UTC')
PPD	Biostatistics Approval	04-Feb-2019 13:59 UTC