STATISTICAL ANALYSIS PLAN

PEACE (Pegzilarginase Effect on Arginase 1 Deficiency Clinical Endpoints): A Randomized, Double-Blind, Placebo-Controlled Phase 3 Study of the Efficacy and Safety of Pegzilarginase in Children and Adults with Arginase 1 Deficiency

Protocol Number: **CAEB1102-300A**

Protocol Version: **6.0**

Protocol Date: **11DEC2020**

SAP Version: **3.0**

SAP Date: **11 NOV2021**

APPROVAL SIGNATURE PAGE

Protocol Title: PEACE (Pegzilarginase Effect on Arginase 1 Deficiency Clinical Endpoints): A Randomized, Double-Blind, Placebo-Controlled Phase 3 Study of the Efficacy and Safety of Pegzilarginase in Children and Adults with Arginase 1 Deficiency

Protocol Number: CAEB1102-300A

SAP Date / Version: Version 3.0/ 11NOV2021

Author:

PPD

Aeglea BioTherapeutics, Inc.

Sponsor Approval

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study, are in accordance with the study objectives, and are consistent with the statistical methodology described in the protocol and the clinical development plan.

PPD _____ 14/11/2021

Date

Aeglea BioTherapeutics, Inc.

PPD _____ 11/12/2021

Date

Aeglea BioTherapeutics, Inc.

PPD _____ 11/12/2021

Date

Aeglea BioTherapeutics, Inc.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. REVISION HISTORY

- SAP Version 1.0 – 11DEC2020
- SAP Version 2.0 – 27JUL2021
- SAP Version 3.0 – 12NOV2021

Changes to the analysis plan are detailed in Section 22.

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
2MWT	2-Minute Walk Test
ADaM	Analysis Data Model
ADA	anti-drug antibody
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
ARG1-D	arginase 1 deficiency
AST	aspartate aminotransferase
ATC	anatomical therapeutic chemical
BMI	body mass index
BP	blood pressure
BSID-III	Bayley Scales of Infant and Toddler Development III
CDC	Centers for Disease Control and Prevention
CI	confidence interval
COA	clinical outcome assessment
COVID-19	coronavirus disease 2019
CRO	contract research organization
CSR	clinical study report
DB	double-blind
DNA	deoxyribonucleic acid
EAA	essential amino acid
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
eCRF	electronic Case Report Form
eDC	electronic data capture
EEG	electroencephalogram
FAS	Full Analysis Set
FMA	Functional Mobility Assessment
FMS	Functional Mobility Scale
GC	guanidino compound
GFAQ	Gillette Functional Assessment Questionnaire
GMFCS	Gross Motor Function Classification System
GMFM-D	Gross Motor Function Measure-88 Part D
GMFM-E	Gross Motor Function Measure-88 Part E
HLT	higher-level term

Abbreviation	Definition
ICH	International Council for Harmonisation of the Technical Requirements for Pharmaceuticals for Human Use
IQ	intelligence quotient
ISR	injection site reaction
IV	intravenous
IXRS	interactive web/voice response system
LFT	liver function test
LS	least squares
LTE	long-term extension
MAR	missing at random
MAS	Modified Ashworth Scale
MCT	meaningful change threshold
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model for repeated measures
NIH	National Institutes of Health
PAP	psychometric analysis plan
PD	pharmacodynamic
PDF	Portable Document Format
PedsQL	Pediatric Quality of Life
PK	pharmacokinetic
PP	per protocol
PT	preferred term
QoL	quality of life
QTc	QT interval corrected
QW	once weekly
RBC	red blood cell
RTF	Rich Text Format
SAE	serious adverse event
SAP	Statistical Analysis Plan
SC	subcutaneous
SD	standard deviation
SDC	Statistics & Data Corporation
SDTM	Study Data Tabulation Model
SF-36	36-Item Short Form Health Survey
SMQ	standardized MedDRA query
SOC	system organ class
SRC	Safety Review Committee
TEAE	treatment-emergent adverse event

Abbreviation	Definition
TFLs	tables, figures, and listings
ULN	upper limit of normal
US	United States
VABS-II	Vineland Adaptive Behavior Scales-II
WAIS-IV	Wechsler Adult Intelligence Scale IV
WHO	World Health Organization
WISC-V	Wechsler Intelligence Scale for Children V
WPPSI-IV	Wechsler Preschool and Primary Scale of Intelligence IV
ZBI-12	Zarit Burden Interview (12-item)

3. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and reporting for protocol CAEB1102-300A version 6.0 dated 11 December 2020.

This SAP is being written with due consideration of the recommendations outlined in the most recent International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9 R1 Guidance for Industry: Statistical Principles for Clinical Trials and the most recent ICH E3 Guidance for Industry: Structure and Content of Clinical Study Reports.

This SAP describes the data that will be analyzed, the specific statistical methods used, and the subject characteristics and efficacy and safety assessments that will be evaluated. The statistical analysis methods presented in this document will supersede the statistical analysis methods described in the clinical study protocol or previous versions of the SAP. If additional analyses are required to supplement the planned analyses described in this SAP, they may be completed and will be identified as post-hoc in the clinical study report (CSR). If a future protocol amendment necessitates a substantial change to the statistical analysis of the study data, this SAP will be amended accordingly.

The analyses will include the following 2 analyses based on the data cut of 14OCT2021:

1. Double-Blind Period: After all subjects have completed 24 weeks (Study Day 162/Visit ID 24) or have discontinued early (hereafter referred to as the double-blind [DB] analysis). Only data through 24 weeks will be included in this analysis.
2. LTE: After all subjects have completed 24 weeks (Study Day 162/Visit ID 24) and started the LTE, select safety and efficacy analysis will be conducted.

4. STUDY OBJECTIVES AND ENDPOINTS

This multi-center, randomized, DB, placebo-controlled Phase 3 study is designed to assess the efficacy and safety of pegzilarginase in children and adults with arginase 1 deficiency (ARG1-D).

4.1. Primary

The primary objective is to demonstrate the efficacy of pegzilarginase relative to placebo based on a statistically significant decrease in plasma arginine levels.

The primary endpoint is as follows:

- Change from Baseline in plasma arginine after 24 weeks of study treatment.

4.2. Key Secondary

The key secondary objective is to demonstrate the efficacy of pegzilarginase relative to placebo based on measures of mobility.

There are 2 key secondary endpoints as follows:

- Mean change from Baseline at Week 24 in the 2-Minute Walk Test (2MWT), and
- Mean change from Baseline at Week 24 in the Gross Motor Function Measure-88 Part E (GMFM-E).

4.3. Other Secondary

Other secondary objectives and endpoints for this study are as follows:

- To compare pegzilarginase with placebo with respect to the proportion of subjects whose endpoint arginine value falls below target guidance of 200 μM. For this objective, a subject is considered a responder if their endpoint arginine value is <200 μM after 24 weeks of study treatment; otherwise, the subject is considered a non-responder.
- To compare pegzilarginase with placebo with respect to the proportion of subjects whose endpoint arginine value falls within the normal range of 40 to 115 μM. For this objective, a subject is considered a responder if their endpoint arginine value is in the normal range (≥40 μM and ≤115 μM) after 24 weeks of study treatment; otherwise, the subject is considered a non-responder.
- To compare pegzilarginase with placebo for changes in ornithine and guanidino compounds (GCs). The endpoints are as follows:
  - Change from Baseline in ornithine and the following GCs after 24 weeks of study treatment:
    - ARGA – Argininic acid
    - GAA – Guanidinoacetic acid
    - GVA – alpha-keto-d-guanidinovaleric acid

    - NAArg – alpha-N-acetylarginine

- To compare pegzilarginase with placebo with respect to other aspects of mobility. There are 5 endpoints, as follows:
  - Mean change from Baseline at Week 24 in the GMFM-88 Part D (GMFM-D).
  - Mean change from Baseline at Week 24 in the following 3 Functional Mobility Scale (FMS) endpoints:
    - FMS 5
    - FMS 50
    - FMS 500
  - Mean change from Baseline at Week 24 in the Gillette Functional Assessment Questionnaire (GFAQ).
- To compare pegzilarginase with placebo with respect to adaptive behavior. There are 5 endpoints from the Vineland Adaptive Behavior Scale (VABS)-II that will be used.
  - Mean change from Baseline in the Adaptive Behavior Composite at Week 24
  - Mean change from Baseline in the Communication Domain at Week 24
  - Mean change from Baseline in the Daily Living Skills Domain at Week 24
  - Mean change from Baseline in the Socialization Domain at Week 24
  - Mean change from Baseline in the Motor Skills Domain at Week 24
- To evaluate the safety and immunogenicity of pegzilarginase. Adverse events (AEs) and anti-drug antibodies (ADAs) will be analyzed.
- To further characterize the pharmacokinetic (PK) profile of pegzilarginase, PK data will be analyzed.

4.4. Tertiary

The tertiary objectives and endpoints for this study are as follows:

- To compare the proportion of responders between pegzilarginase and placebo in the composite clinical outcome. Response for the composite clinical outcome is defined in Section 18.1 of the SAP using the 2MWT, GMFM-D, and GMFM-E.
- To compare the proportion of responders between pegzilarginase and placebo with respect to other aspects of mobility. Response for FMS and GFAQ is defined as a 1 -level change at any distance and a 2-level change at any distance, respectively.
- To compare the proportion of responders between pegzilarginase and placebo with respect to adaptive behavior. Response is defined as improvement by ≥ 7.5 points within the VABS-II for each of the following: the overall composite and the communication, daily living skills, socialization, and motor skills domains.

- To compare pegzilarginase to placebo with respect to improvement of spasticity and objective measures of neurological/neuromotor manifestations. The endpoint is the change from Baseline at Week 24 in the Modified Ashworth Scale (MAS).
- To further describe caregiver and clinician global impression of the impact of pegzilarginase on motor function as measured by the Caregiver/Clinician Global Impression of Change and Caregiver/Clinician Global Impression of Severity.
- To describe the impact of pegzilarginase on fine motor function using the 9-Hole Pegboard Test.
- To describe the impact of pegzilarginase on quality of life (QoL) using the Pediatric Quality of Life (PedsQL), 36-Item Short Form Health Survey (SF-36), and the Zarit Burden Interview (12-item) (ZBI-12).
- To describe the impact of pegzilarginase on neurocognitive functioning and memory using the Bayley Scales of Infant Development (BSID)-III (as age appropriate) and Wechsler Intelligence Batteries.
- To describe diets maintained by the subjects in the study using 3-day diet records.
- To describe the effect of pegzilarginase on plasma ammonia levels.
- To evaluate the effects of pegzilarginase on growth in pediatric subjects using Z--scores for height, weight, and body mass index (BMI). The Z-scores will be computed using the Centers for Disease Control and Prevention (CDC) growth curves for subjects younger than 18 years of age.
- To characterize the long-term (>24 to approximately 178 weeks) treatment effects of pegzilarginase administration once weekly (QW), in particular, to evaluate the extent to which efficacy is sustained over the LTE through 48 weeks (i.e., LTE24).

5. STUDY DESIGN AND PROCEDURES

5.1. General Study Design

This is a Phase 3, multi-center, randomized, DB, placebo-controlled study to evaluate the efficacy and safety of pegzilarginase in subjects with ARG1-D. To be considered for this study, male and female subjects aged ≥2 years must provide written informed consent/assent and have a diagnosis of ARG1-D as documented by medical records, which must include 1 of the following: elevated plasma arginine levels, a mutation analysis that results in a pathogenic variant, or red blood cell (RBC) arginase activity.

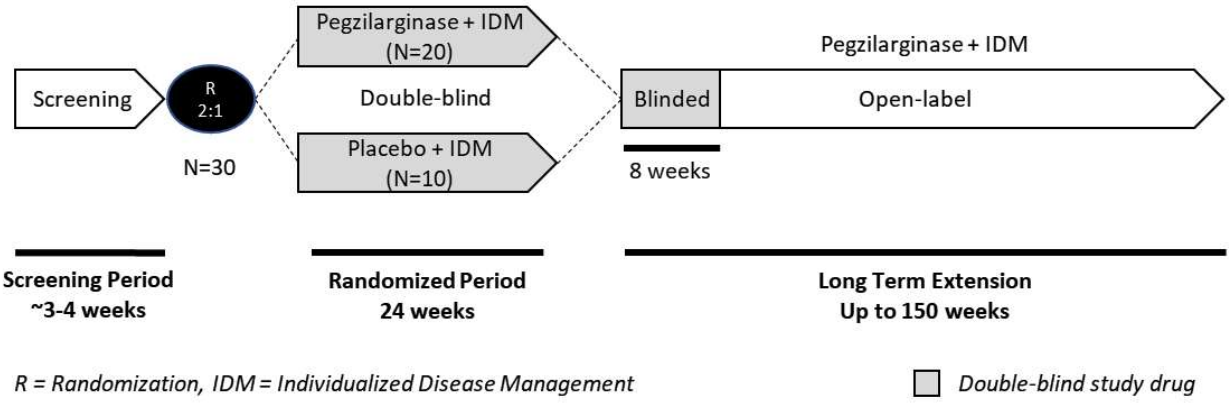
This study design, as shown in Figure 1, will consist of:

- Screening Period: Lasting 3 to 4 weeks, this period will include collection of all necessary information to ensure eligibility, establish baseline plasma arginine measures, collect diet data, and determine adherence to a prescribed diet as reported in a diet diary.
- DB Randomized Period: Lasting 24 weeks.
- LTE Period: Lasting approximately 150 weeks in which all subjects will receive pegzilarginase. The first 8 weeks of this period will remain blinded to ensure that study data relating to the randomized period is collected prior to unblinding.

Subjects will be randomized to treatment following completion of all screening assessments and confirmation of study eligibility in a 2:1 ratio to receive weekly intravenous (IV) infusion of pegzilarginase or placebo during the 24-week DB period. Subjects should continue their previously prescribed protein restriction in addition to any Essential Amino Acid (EAA) supplementation and/or ammonia scavengers.

Following the DB period, each subject will enter the LTE period where, for the first 8 weeks, all subjects will continue to receive a blinded dose of study drug. No placebo will be administered in the LTE period; thus, those taking placebo in the DB period will transition to pegzilarginase in the LTE period where the first 8 weeks will be blinded. Following the first 8 weeks in the LTE period, subjects will have the option to receive pegzilarginase by subcutaneous (SC) injection with investigator and sponsor approval or continue with IV dosing.

Figure 1: Study Design/Schema



5.2. Schedule of Assessments

Assessments are to be performed as outlined in the protocol. For the purpose of defining visits in this SAP, Visit Weeks will refer to corresponding Visit IDs defined in the protocol Schedule of Assessments (see Protocol Section 1.2).

6. STUDY TREATMENTS

Subjects will be administered either pegzilarginase or placebo during the DB period. All subjects will receive pegzilarginase during the LTE period of the study. The dose of drug to be administered will be monitored at each site by an unblinded pharmacist, and an unblinded physician where required, to manage protocol-defined dose adjustments.

In addition to study drug, subjects will also continue their individualized protein restricted diet and any supplemental Essential Amino Acids and/or ammonia scavengers.

6.1. Method of Assigning Subjects to Treatment Groups

Subjects who pass all screening criteria will be randomized to either pegzilarginase or placebo in a 2:1 ratio. Randomization in this study is stratified by severity of prior history of hyperammonemia. The purpose of this stratification is to avoid potential imbalance in this factor across randomized treatment groups.

The randomization strata, based on hyperammonemic episodes and implemented by the interactive web/voice response system (IXRS), are:

- ≥1 hyperammonemic episode within 90 days of consent or ≥2 hyperammonemic episodes within 365 days of consent
- No hyperammonemic episodes within 365 days of consent, or 1 hyperammonemic episode occurring >90 days prior to consent

6.2. Blinding and Unblinding

Laboratory results for arginine and ornithine have the potential to unblind the investigator to the subject's treatment group. Therefore, results for these laboratory tests performed during the 24 -week DB period will not be provided to the investigator or other blinded individuals, including subjects, families, sponsor personnel, or assessors, until subjects have completed the blinded portion of the study to LTE01 and the database to this timepoint has been frozen and formal unblinding of the 24-week DB period has occurred. A subject is only considered formally unblinded when his/her DB randomized treatment has been provided to the investigator or the subject. Inadvertent exposure to blinded data such as arginine level is not considered a formal unblinding of the subject.

Methods for emergency unblinding are provided in the IXRS manual.

7. SAMPLE SIZE AND POWER CONSIDERATIONS

Based on data from Study 101A and Study 102A, arginine levels at 24 weeks post-baseline were reduced to -2.13 on the $\log_2$ scale, representing at 77% decrease from baseline. The $\log_2$ scale standard deviation (SD) was estimated at 0.681. Given that it is unlikely that placebo will have any true treatment effect on arginine levels at 24 weeks, the reduction at 24 weeks for placebo is estimated to be 0.

Using these assumptions, a total sample size of 30 subjects, with 20 subjects assigned to the pegzilarginase treatment group and 10 subjects assigned to the placebo treatment group, will provide over 95% power to detect a statistically significant difference in the reduction of arginine levels at the 2-sided α=0.05 level of significance. Additionally, the power to test for a true treatment effect on arginine inhibition between pegzilarginase and placebo is outlined below in Table 1 for varying theoretical degrees of inhibition of placebo with N=30 subjects in a 2:1 ratio.

Table 1: Power Calculations for Varying Estimates of True Treatment Difference

True Effects on Arginine at 24 Weeks			Power at 1-Sided α=0.025
Pegzilarginase	**Placebo**	**Difference**	
-75%	0%	-75%	>99%
-75%	-5%	-70%	>99%
-75%	-10%	-65%	>99%
-75%	-15%	-60%	>99%
-75%	-20%	-55%	>99%
-75%	-25%	-50%	>99%
-75%	-30%	-45%	>99%
-75%	-35%	-40%	>99%
-75%	-40%	-35%	>99%
-75%	-45%	-30%	98%
-75%	-50%	-25%	95%

Power calculations have also been performed for the 2 key secondary endpoints using simulated data as follows.

- Baseline and 24-week data were sampled from a joint distribution of the 2MWT and GMFM-E based on assumed effect sizes seen in Studies 101A and 102A.
- Subjects were classified according to the Gross Motor Function Classification System (GMFCS), which assigns gross motor function capabilities based on movements such as sitting, walking and use of mobility devices with 5 categories ranging from I (most functional) to V (transported in wheelchair in all settings).

As shown in Table 2, the proportion of subjects used in the simulations in Class I was 50% and the proportion in Class II and III was therefore also 50%.

Table 2: Key Secondary Endpoints Mean and Variance Assumptions Used in Power Simulations

	Class I GMFCS	Class II and III GMFCS
2MWT Baseline mean (SD)	134.9 (37.1)	60.2 (18.8)
2MWT Week 24 mean (SD)	146.6 (40.6)	62.0 (20.1)
GMFM-E Baseline mean (SD)	69.1 (3.0)	23.0 (9.4)
GMFM-E Week 24 mean (SD)	70.3 (2.9)	32.2 (10.6)

Placebo subjects were assumed to have Week 24 endpoint values with the same mean and variance as their Baseline values (i.e., zero mean change from Baseline).

The correlation matrix used for sampled data is detailed in Table 3 below. To be conservative, the values shown in the table are rounded down from the actual values generated from the simulation.

Table 3: Key Secondary Endpoints Correlation Assumptions Used in Power Simulations

	2MWT Baseline	2MWT Week 24	GMFM-E Baseline	GMFM-E Week 24
2MWT Baseline	1	0.9	0.8	0.8
2MWT Week 24		1	0.8	0.8
GMFM-E Baseline			1	0.9
GMFM-E Week 24				1

An analysis of covariance (ANCOVA) was performed on the change from Baseline for each endpoint and using a Hochberg multiplicity adjustment to control the global Type 1 error of the study, it was shown that there was more than 90% power to show statistical significance in at least 1 of the key secondary endpoints.

Although an exact screen failure rate cannot be accurately predicted based on available data, a sufficient number of subjects will be screened so that at least 30 subjects in the Full Analysis Set (FAS) have at least 1 follow-up measurement on which clinical response can be assessed (2MWT and GMFM-E).

Subjects who discontinue after their first dose but before completing the DB portion (“dropouts”) will not be replaced.

8. DATA PREPARATION

8.1. Input Data

Study data will be collected by independent vendors and sent to the contract research organization (CRO) who will be managing the data for this study. The CRO will be responsible for providing coded values for medical history, AEs, and prior and concomitant procedures using MedDRA Version 23.1 or later. Prior and concomitant medications will be coded by the CRO using the World Health Organization (WHO) Drug Dictionary version dated June 2016 or later.

Data transfer will be based on study deliverable timelines determined by all parties involved.

8.2. Output Data

Data from the data provider will be transferred and incorporated into standard formats following the Study Data Tabulation Model (SDTM). Data will then be mapped to analysis datasets using the Analysis Data Model (ADaM). Both SDTM- and ADaM-formatted data will be used to create the tables, figures, and listings (TFLs).

SDTM will follow the SDTM version 1.4 model and will be implemented using the SDTM Implementation Guide version 3.2 and the SDTM Controlled Terminology version 2016-06-24. ADaM data will follow the ADaM version 2.1 model and will be implemented using the ADaM Implementation Guide version 1.3. Both SDTM and ADaM will be validated using Pinnacle 21 version 2.2. Any discrepancies in the validation will be noted in reviewer's guides accompanying the final data transfers.

Define.xml will be created for SDTM and ADaM using the Define-XML version 2.0 model.

Note: if during the course of this study there are any updates in versioning to the above criteria, the most recent version may be utilized.

9. ANALYSIS PERIODS AND POPULATIONS

9.1. Analysis Periods

For study population and efficacy outputs comparing pegzilarginase vs placebo, subjects will be categorized according to the treatment to which they were randomized.

For safety analyses, subjects will be categorized according to the treatment which they actually received. Subjects receiving a single dose of pegzilarginase will be categorized in the pegzilarginase treatment group for all safety outputs.

For selected outputs, data summaries and analyses will be provided for the following periods.

9.1.1. Placebo-Controlled DB Period – Baseline to Week 24

This period starts on the day of randomization (Study Day 1/Visit 1) and stops on:

- The LTE Study Day 1/Visit LTE01; or
- The day of the discontinuation visit if the subject discontinued prior to or at LTE Study Day 1/Visit LTE01; or
- The last known study day if the subject is lost to follow-up prior to LTE Study Day 1/Visit LTE01 during the period.

Analysis:

- **Efficacy and Safety**: The primary, key secondary, secondary, tertiary, and safety endpoints will be summarized for this analysis period.

9.1.2. LTE Period – LTE01 subgroup to LTE150 (End of Study)

- The LTE Period starts on: LTE Study Day 1/Visit LTE01.

The LTE Period stops on:

- The data cut-off date; or
- The day of the discontinuation visit, if the subject discontinued prior to or on LTE Study Day 1044/Visit LTE150; or
- The last known study day if the subject was lost to follow-up prior to LTE Study Day 1044/Visit LTE150; or
- LTE Study Day 1044/Visit LTE150 visit date (i.e., the last scheduled study visit).

Comparisons during this analysis period will be by treatment regimen (i.e., placebo - pegzilarginase (pbo-peg) vs pegzilarginase - pegzilarginase (peg-peg)).

Analysis:

- **Efficacy data** (primary, key secondary) will be presented through the data cutoff, in tables or listings.
- **Safety data** (i.e., AEs, labs, vitals, electrocardiograms [ECGs]) will be presented through the data cutoff in tables or listings.

9.2. Analysis Populations

9.2.1. Consented Set

The Consented set includes all subjects who sign an informed consent form. This population will be used to report screening data.

9.2.2. Randomized Set

The Randomized set includes all subjects in the Consented set who are randomized to a blinded study treatment.

9.2.3. Full Analysis Set

The FAS includes all subjects who are randomized and who receive at least 1 dose of blinded study treatment. All safety and efficacy analyses will be performed on this set.

9.2.4. Per-Protocol Set

The Per-Protocol (PP) set includes all subjects in the FAS who are compliant with the protocol without major protocol deviations and have adequate exposure as defined below during the DB period that directly impact efficacy analyses. These include the following:

- Subjects who met all inclusion/exclusion criteria
- Subjects who were assigned into the correct randomization strata
- Subjects who completed the treatment as originally allocated
- Subjects who have not missed either >2 doses or 2 consecutive doses in last 4 weeks (Weeks 21, 22, 23, 24)
- Subjects who are at least 80% compliant with dosing
- Subjects who have no important protocol deviations that directly impact the results of the primary or key secondary analyses. Important protocol deviations that could exclude subjects from the PP set include but are not limited to the following:
  - Receiving prohibited concomitant medications or significant changes in concomitant medications (i.e., non-stable use; see protocol based on clinical review)
  - Surgeries to correct disease-related abnormalities
  - Significant changes in Total protein (EAA/natural) that, based on clinical review, would impact efficacy

The final determination of these violations, and thereby the composition of the PP population, will be determined and separately documented prior to the unblinding of the 24-week DB period. For the analyses using the PP set, subjects will be included in the treatment group to which they were randomized.

9.2.5. Pharmacokinetic Set

The PK set includes all subjects receiving at least 1 dose of pegzilarginase who have sufficient data to calculate PK parameters for pegzilarginase. Further details of the PK, Pharmacodynamic (PD), PK/PD, and Immunogenicity sets and methods for their analyses will be presented in a separate PK SAP.

10. GENERAL STATISTICAL CONSIDERATIONS

10.1. Missing or Inconclusive Data Handling

For the primary efficacy endpoint analysis (plasma arginine) and the secondary endpoint of GCs and ornithine, when a final value as per Section 17.1 is not available, the change from Baseline will be imputed as zero. This imputation is detailed for each of the relevant endpoints in Section 17 (efficacy analysis). Missing data due to the following intercurrent events: withdrawal from study, death, and coronavirus disease 2019 (COVID-19) will therefore be imputed as though the subject does not improve from Baseline: a composite estimand strategy.

Other key secondary and secondary endpoints will not be imputed. Analyses for key secondary endpoints are mixed models for repeated measures (MMRMs), which produce estimates that are consistent with a missing at random (MAR) assumption for missing data.

Partial/missing start and end dates for AEs, concomitant medications, and medical history will be imputed as follows so that a complete date can be used to flag data as treatment emergent or concomitant with treatment:

Partial/missing start date:

- Dates with missing day only will be imputed as the 1st of the month unless the month and year are same as the month and year of first dose of study medication, in which case missing day will be imputed as the first dose day of study medication.
- Dates with both day and month missing will be imputed as 1 Jan unless the year is same as the year of first dose of study medication, in which case missing day and month will be imputed as the first dose day and month of study medication.
- Completely missing dates will be imputed as the first dose date of study medication unless the end date is on or before the first dose date of study medication, in which case missing date will be imputed as 1 Jan of the same year as the end date.
- In the event a result is negative the value will be imputed as 0.

Partial/missing end date:

- Dates with missing day only will be imputed as the last day of the month unless the month and year are the same as the month and year of the last dose of study medication, in which case missing day will be imputed as the last dose day of study medication.
- Dates with both day and month missing will be imputed as 31 Dec unless the year is same as the year of the last dose of study medication, in which case missing day and month will be imputed as the last dose day and month of study medication.
- If the ongoing flag is missing or “Yes,” then the date will not be imputed unless death date is available, in which case the missing date will be imputed as the death date. If ongoing is “No” then the missing end date will be imputed as the last dose date.
- If the imputed date is after the date of death, then the end date will be set equal to the date of death.

The original dates will be displayed in data listings and the imputed dates will be used in derivations only (study day, treatment-emergence status, etc.).

10.2. Handling of Assessments from Re-Starting Study Due to COVID-19 Pause

Subjects may re-start the study if they have had a temporary pause in dosing due to the COVID-19 pandemic. All subjects paused due to COVID-19 were allowed to re-start participation in accordance with the Study Guidance for COVID-19 Subject Handling Document.

Dependent on the duration of the pause and study stage, baseline neuromotor and neurocognitive assessments were repeated (described in individual subject visit continuation plans) and used as baseline assessments for study analysis according to the criteria in Table 4 below.

Table 4: Main Criteria for Repeating Assessments

Type of Assessment	Criteria	Action
Neuromotor	Pause ≥8 weeks	Re-assess 2MWT and GMFM-D and -E
	Pause <8 weeks and change in • GMFCS of 1 point or • FMS 5; 50; 500 of 1 point or • GFAQ of 2 points	Re-assess 2MWT and GMFM-D and -E
Neurocognitive	Pause >4 weeks	Re-assess PedsQL, SF-36, ZBI-12
	Pause >12 weeks	Re-assess VABS
	Pause >24 weeks	Re-assess BSID-III, Wechsler scales re-start will have a new set of Baseline values defined using the last measurements prior to the new Visit 1 (first dose of study drug after the re-start)

10.3. Safety Data

Study sites continued to monitor subjects during the pause due to COVID-19. All AEs reported during the pause will be included in the listings.

Safety assessments (labs, vitals, etc.) were conducted according to protocol for the visits conducted.

10.4. Definition of Baseline

For the DB period, Baseline for summary of baseline disease characteristics is defined as the last value measured on or before the day of the first dose of study drug, just prior to dosing.

10.4.1. Baseline Handling for Study Pauses and Re-Starts Due to COVID-19

10.4.1.1. Definition of Baseline for Subjects Who Paused During Screening

Subjects who paused during screening due to COVID-19 were allowed to re-start the study. At study re-start, the site continued screening the subject, repeating assessments as necessary, conducted in up to 3 further visits following the original protocol for screening as per Table 4. These visits were assigned as screening visits or unscheduled visits according to the COVID-19 data handling memo. Previous baseline assessments were reused (e.g., medical history) or updated if necessary. Dependent on the duration of the pause baseline neuromotor and neurocognition assessments were repeated according to the criteria in Table 4.

For the efficacy analysis, if there are 2 assessments at Baseline for a subject who paused the assessment, the later screening visit or unscheduled visit will be used. If only 1 baseline assessment was performed, that one will be used in the analysis.

10.4.1.2. Definition of Baseline for Subjects Who Paused During the Placebo-Controlled DB Period (Week 1-Week 24)

Subjects who paused during the DB period due to COVID-19 were allowed to re-start the study in accordance with the Study Guidance for COVID-19 Subject Handling Document. Subjects who had received fewer than 6 doses of study drug were re-started from Dose 1 and repeated previous visits using unscheduled visit forms. Subjects who had received 6 or more doses were to be handled on an individual basis. At study re-start, the site repeated baseline assessments in up to 3 further visits; these visits were assigned as unscheduled visits according to the COVID-19 data handling memo. Previous baseline assessments were reused (e.g., medical history) or updated if necessary. Dependent on the duration of the pause, baseline neuromotor and neurocognition assessments were repeated according to the criteria in Table 4. All collected subject data will be included in the listings.

If the subjects were re-started from Dose 1, the re-start baseline assessments will be used for the efficacy analysis, even though the assessment was taken after the first dose of treatment; this is deemed acceptable under the circumstances created by the COVID-19 pandemic and due to the low number of prior doses received followed by a significant washout period (22 to 33 weeks). If there are 2 or more assessments at Baseline for a subject who paused during dosing, the assessment from the later baseline visit or unscheduled visit will be used. If only 1 baseline assessment was performed, that one will be used in the analysis.

10.4.1.3. Definition of Baseline for Subjects Who Paused During the LTE Period

Subjects who paused during the LTE period due to COVID-19 were allowed to resume the study in accordance with the Study Guidance for COVID-19 Subject Handling Document. Subjects re-started dosing at the visit that was due to be conducted next when the pause occurred. There was no repeat of baseline assessments prior to re-starting dosing in these subjects.

10.4.2. Baseline for Plasma Arginine, Ornithine, and GCs

For the primary endpoint, change from Baseline in plasma arginine after 24 weeks of study treatment, baseline plasma arginine is defined as the arithmetic mean of all plasma arginine values (analyzed at the designated central laboratory) obtained during the screening period and

prior to the first dose of blinded study treatment. If a subject is re-screened, the latest value taken during the final screening visit will be used in the computation of baseline plasma arginine.

The same derivation for plasma arginine will be utilized for calculating the baseline values for ornithine and GCs.

10.4.3. Analysis Visit Windows for LTE Period

To ensure that data for the clinical outcomes is collected at proper visits during the LTE period, a window of ± 3 visits will be applied to the appropriate target visits where the data is collected (approximately every 12 weeks as per the study schedule of events). The scheduled visits with windows is below:

LTE Visit	Visit Window	days
LTE12	LTE09 – LTE15	57 - 99
LTE24	LTE021– LTE27	141 - 183
LTE36	LTE33 – LTE39	225 - 267
LTE48	LTE45 – LTE51	309 - 351
LTE60	LTE57 – LTE63	393 - 435
LTE72	LTE69 – LTE75	477 - 519
LTE84	LTE81 – LTE87	561 - 603
LTE96	LTE93 – LTE99	645 - 687
LTE120	LTE117 – LTE123	813 - 855
LTE150	LTE147 – LTE150	1023 - 1044

10.5. Data Analysis Conventions

Statistical programming and analyses will be performed using SAS® version 9.4 or higher. Output will be provided in rich text format (RTF) for tables and both RTF and portable document format (PDF) for all TFLs using landscape orientation.

Summaries for continuous and ordinal variables will include the number of observations (n), arithmetic mean, SD, median, minimum, and maximum. Minima and maxima will be reported with the same precision as the raw values; means and medians will be presented to 1 additional decimal place than reported in the raw values. Standard deviations will be presented to 2 additional decimal places than reported in the raw values. Summaries for discrete variables will include counts and percentages. All percentages will be rounded to 1 decimal place (i.e., xx.x%). Differences between pegzilarginase and placebo will be calculated as pegzilarginase minus placebo and change from Baseline will be calculated as post-Baseline visit minus Baseline, unless otherwise specified. Results (least squares [LS] means, LS mean differences, and their 95% confidence intervals [CIs]) from analyses performed on logged data will be exponentiated prior to presentation. The exponentiated treatment group LS mean from the analysis of the final value will represent a treatment group geometric mean. The exponentiated treatment group LS mean from the analysis of the change from Baseline will represent a treatment group ratio to Baseline. The exponentiated treatment group LS mean difference will

represent the treatment effect in terms of a relative ratio: pegzilarginase/placebo. These ratios can be converted to percentage changes using the formula: percentage change = 100*(1-ratio).

All statistical tests will be 2-sided with a significance level of 0.05 (α=0.05), unless otherwise specified. CIs for differences between treatment groups will be 2-sided at 95% confidence. All p-values will be rounded to 4 decimal places; p-values less than 0.0001 will be presented as "<0.0001", and p-values greater than 0.9999 will be presented as ">0.9999."

Unless otherwise specified, summaries will be presented by treatment group and, where appropriate, visit. Listings will be sorted by treatment group, subject number, visit/timepoint, and parameter as applicable.

10.6. Subgroup Analyses

The following subgroups will be defined for any subgroup analyses.

- Age (<18 years old at screening, ≥18 years old at screening)
- Sex (Male, Female)
- Region (US, non-US)
- GMFCS Classification at Baseline (I, ≥II)

Subgroup analyses will be performed on the primary analysis of change from Baseline in plasma arginine after 24 weeks of study treatment, and the key secondary analyses of change from Baseline at Week 24 in the 2MWT and change from Baseline at Week 24 in GMFM-E if numbers within the subgroups are sufficient (i.e., ≥4 subjects in each group).

10.7. Adjustments for Multiplicity and Global Control of Type 1 Error

Nominal p-values may be presented on all outputs, however the following strategy, detailed in Table 5, is to be used in order to declare formal statistical significance of certain endpoints while also controlling the global Type 1 error of the study.

Table 5: Methods Used for Global Control of Type 1 Error

Type of Analysis	Analysis Set	Endpoint	Type 1 Error Control/Multiplicity Adjustment
Primary Analysis	FAS	Comparison of change from Baseline in logged plasma arginine after 24 weeks of study treatment.	The significance test of the primary endpoint analysis will be declared statistically significant if the p-value (2-sided) is ≤ 0.05.
Formal testing of the key secondary endpoints will proceed only if the significance test of the primary endpoint analysis is declared statistically significant.			

Type of Analysis	Analysis Set	Endpoint	Type 1 Error Control/Multiplicity Adjustment
Key Secondary Analyses	FAS	a) Comparison of mean change from Baseline at Week 24 in the 2MWT. b) Comparison of change from Baseline at Week 24 in the GMFM-E.	The Hochberg procedure (Hochberg 1988) will be used to determine the formal statistical significance of one or both key secondary endpoint analysis comparisons. The procedure is as follows: • If both p-values (2-sided) are ≤0.05 then both endpoints will be declared statistically significant. • If not, then if the smallest p-value (2-sided) is ≤0.025, then that endpoint alone will be declared statistically significant.
Formal testing of the remaining secondary endpoints will proceed only if both key secondary endpoints are statistically significant.			
Additional Secondary Analyses	FAS	Comparison of response (arginine <200 µM after 24 weeks of study treatment).	The significance test of the next secondary endpoint analysis will be declared statistically significant if the p--value (2-sided) ≤ 0.05 (or ≤ 0.025 if only one of the above-defined key secondary endpoints is statistically significant.)
Formal testing of the next secondary endpoint will proceed only if the significance test of the previous secondary endpoint analysis is declared statistically significant.			
Additional Secondary Analyses	FAS	Comparison of response (arginine 40 to 115 µM after 24 weeks of study treatment).	The significance test of the next secondary endpoint analysis will be declared statistically significant if the p-value (2-sided) ≤ 0.05 (or ≤ 0.025 if only one of the above-defined key secondary endpoints is statistically significant.)
Formal testing of the next secondary endpoint will proceed only if the significance test of the previous secondary endpoint analysis is declared statistically significant.			
Additional Secondary Analyses	FAS	Comparison of mean change from Baseline in the GMFM-D at Week 24.	The significance test of the next secondary endpoint analysis will be declared statistically significant if the p-value (2-sided) ≤0.05 (or ≤ 0.025 if only one of the above-defined key secondary endpoints is statistically significant.)

10.8. Timepoints

Timepoints in all analyses are based on the visits/study weeks as recorded in the database. Appropriate dates (e.g., date of measurement or date of specimen collection) will first be used to ensure only data (including data from scheduled, unscheduled, and discontinuation) measured or collected within the specific analysis period are included, and then the visits/study weeks as recorded in the database will be used to assign 1 value (possibly missing) to each timepoint at the subject level according to the rules presented in the subsequent sections.

10.8.1. Scheduled Visits

If a value is available at a scheduled visit, then that value will be used for the study visit, regardless of whether a value is available from an unscheduled visit that is associated with the scheduled visit.

If the value at a scheduled visit is missing but a value is available at an associated unscheduled visit (within the same analysis period), the value from the associated unscheduled visit will be used for the study visit. If values from more than 1 associated unscheduled visit (e.g., Visits 8.01

and 8.02) are available, then the value from the first associated unscheduled visit (e.g., Visit 8.01) will be used for the study visit.

If the value at a scheduled visit is missing, and there is no value available from an associated unscheduled visit, the value at the study visit will be missing.

Note that unscheduled visits are recorded to 1 general visit code. The visit numbering for unscheduled visits will be re-numbered based on the unscheduled visit date (and time) comparison to the scheduled visit date (and time). If the unscheduled record falls in between 2 scheduled visits, the re-numbered value will be assigned. For example, if the unscheduled visit date is “on or after Visit 1” AND “prior to Visit 2”, then the re-numbered value will be assigned based upon the date between the 2 visits. If there are multiple unscheduled visits in that interval, the unscheduled visit date (and time) will be used to order the re-number values.

Unscheduled visits were also used to incorporate previously scheduled visit data for those subjects who paused during the study due to COVID-19. Please see Sections 10.2 and 10.4.1 regarding the handling of COVID-19 pauses.

10.8.2. Discontinuation Visit

A value from the discontinuation visit (for a discontinued subject) will be mapped to the next scheduled visit after the last visit the subject has completed.

Note that some measurements collected at the discontinuation visit may be mapped to a non-scheduled visit for that measurement. For example, while the SF-36 is not collected at Study Day 43/Visit ID 7, if a subject’s last completed visit is Study Day 36/Visit ID 6, then the discontinuation visit for that subject would be mapped to Study Day 43/Visit ID 7.

10.8.3. End of Specific Analysis Period Visit

The end of the DB period is defined as the occurrence of the LTE01 visit.

The end of the LTE period visit is defined as the last study visit.

11. DISPOSITION OF SUBJECTS

Subject disposition will be reported based on the DB treatment. Completers are defined as those subjects who did not discontinue from the study prior to LTE01; thus, completing the 24-week DB period. The number of subjects screened, the number of subjects who screen failed, and the number and percentage of subjects randomized will be summarized. The failed inclusion/exclusion criteria of subjects who were screened and not randomized will be included in the summary. The above percentages will be based on the number of subjects screened.

The percentages for number of subjects in the FAS and PP sets, as well as the number of subjects who completed the 24-week DB period, discontinued during the 24-week DB period, and discontinued during the LTE period, will be based on the Randomized set.

Disposition will be summarized by treatment group and overall unless otherwise specified. The following subject data will be summarized:

- Number of subjects who were screened for the study (Consented set – overall column only)
- Number of subjects who screen failed (Consented set – overall column only)
- Number (%) of screen failures of the total number of subjects screened for the study
- Number (%) of subjects who were randomized in the DB period (Randomized set)
- Number (%) of subjects in the FAS
- Number (%) of subjects in the PP set
- Number (%) of subjects in the PK set
- Number (%) of subjects who completed the DB period
- Number (%) of subjects who completed the LTE period
- Number (%) of subjects who discontinued during the DB period and the reason for discontinuation
- Number (%) of subjects who discontinued during the LTE period and the reason for discontinuation

Listings of subject enrollment eligibility, analysis sets, treatment assignments, and subject disposition will also be provided.

12. DEMOGRAPHIC AND PRETREATMENT VARIABLES

12.1. Demographic Variables

The following demographic and baseline variables will be summarized at screening and will be summarized by treatment group and overall, for the FAS. Continuous variables will include the following:

- Age (years)
- Weight (kg)
- Height (cm)
- BMI (kg/m^2)

Categorical variables will include the following:

- Age categories (2 - <6, 6 - <12, 12 - <18, ≥18 years)
- Race
- Ethnicity
- Region (US, non-US)
- Sex (Male, Female)

If a subject's date of birth is collected, age will be calculated using the following formula:

Age = (Informed Consent Date – Date of Birth) / 365.25, truncated as an integer

In addition, summary tables will be generated to report Baseline disease characteristics and family medical history.

Subject listings will also be provided on demographic variables for the FAS.

12.2. Baseline Disease Characteristics

Baseline characteristics will be summarized descriptively by treatment group for the FAS and will include the following:

- Age at initial symptoms (years)
  - Disease manifestations at diagnosis OR newborn screening history
- Age at diagnosis (years)
- Method of diagnosis (i.e., arginine levels, RBC enzyme activity, deoxyribonucleic acid [DNA] testing/other, Newborn screening)
- Newborn screening history
- Use of ammonia scavengers
- Use of antispasmodics (e.g., baclofen)
- Use of botulinum toxin–containing medications

- Tendon release or other surgical procedures
- Performance of DNA testing to detect ARG1-D
- Performance of RBC enzyme activity
- Use of speech, physical, occupational, or other therapy
- History of hyperammonemia
- Disease characteristics: including the following:
  - Spasticity
  - Seizures
  - Cognitive delays
  - Language delays
  - Muscle cramps
  - Hyperammonemia frequency and episodic history requiring hospitalization for previous 2 years
  - Liver injury and/or dysfunction
- Family and sibling history of ARG1-D, including the following:
  - Disease manifestations at diagnosis OR newborn screening history
  - Age at diagnosis (years)
  - Method of diagnosis (ie, arginine, RBC enzyme activity, DNA testing/other, Newborn screening)
- Baseline GMFCS level (I, II, III, IV)

For the “age at” definitions, the following equation will be used:

- Age (years) = (Date of Initial symptom[s] – date of birth +1)/ 30.42/12. If the month and year are provided but the day is missing, the missing day is imputed as 01. If only the year is provided, then the missing month and day are imputed as Jan 1 for the calculation. If the year is missing, the subject will be excluded from the analysis. If the imputed initial symptom(s) date is before date of birth, the age will be imputed as 0 years.

13. MEDICAL HISTORY AND CONCOMITANT MEDICATIONS

13.1. Medical History

Medical history data will be collected during the screening period. A summary table and listing of medical history data will be provided. Tables and listings will be reported on the FAS and will include System Organ Class (SOC) and Preferred Term (PT) variables as coded using MedDRA Version 24.0, or later.

13.2. Prior and Concomitant Medications

The start and stop date of each medication will be recorded to the level of the calendar date. "Prior medications" are defined as any medication that has been discontinued prior to the date of randomization. "Concomitant medications" are defined as any medication taken during the study period. If a medication is started prior to randomization it is considered "ongoing" at the time the subject is randomized, this medication will be considered both a prior and concomitant medication.

Prior and concomitant medications will each be summarized. The DB period will be summarized by treatment group. Prior and concomitant medications will be coded using the WHO Drug Dictionary version dated June 2016, or later, and summarized anatomical therapeutic chemical (ATC) 4 classification and preferred name. If the ATC 4 classification is not provided, then the next lowest classification that is provided in the coding dictionary will be used.

Prior and concomitant medications will be summarized using the FAS. Subjects may have more than 1 reported medication per ATC classification. Subjects with more than 1 medication reported within an ATC classification will only be counted once per ATC classification. Percentages will be based on the number of subjects in each treatment group.

A separate by-subject listing of concomitant medications with verbatim and generic names will be reported.

14. DOSING COMPLIANCE AND TREATMENT EXPOSURE

There are 2 classification periods for both dosing compliance and exposure, as follows:

- DB period
- LTE period

Summary statistics and a listing will be provided for both analysis periods. All data will be reported on the FAS.

Summary statistics for treatment duration (weeks), as well as frequency summaries of treatment duration categories (e.g., <4, ≥4 to <8, ≥8 to < 12, ≥12 to <24, ≥24 weeks), will be provided.

14.1. Dosing Compliance for the Placebo-Controlled DB Period

Subjects are expected to receive a weekly infused dose of DB study drug. The number of doses taken and the number of expected doses (i.e., doses planned prior to withdrawal from the study), and the compliance (number of doses taken / number of expected doses) will be summarized. Compliance will also be reported in categories: ≥80%, 60% to <80%, 40% to <60%, 20% to <40%, and 0% to <20%, if there are sufficient frequencies for meaningful summaries. Subjects may re-start the study if they have had a temporary pause due to the COVID-19 pandemic. Compliance for subjects who re-start will only be calculated using the number of expected doses during the re-started study period (defined in Section 10.4.1).

14.2. Dosing Exposure for the Placebo-Controlled DB Period

In general, treatment exposure (in weeks) for each subject will be calculated as follows:

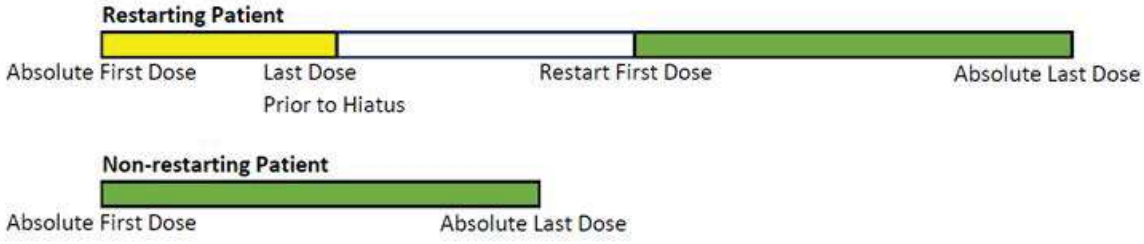
(Date of Last Drug Administration – Date of First Drug Administration +1) /7

The date of last drug administration is defined as the last drug administration prior to and including through Week 24 up to visit LTE01.

The following 4 exposure statistics will be presented, as shown in Figure 2:

- Total duration in study will be calculated as the time from the subject's first visit to the subject's last visit (informed consent to last visit)
- Total exposure to study drug will be calculated using the dates of absolute first and absolute last drug administration per subject minus the time drug was paused, equivalent to yellow plus green periods, minus the white period.
- Efficacy exposure will be calculated using the dates of re-start first dose (where applicable) and absolute last drug administration per subject through 7 days post dosing – equivalent to the green periods only. Subjects who do not re-start will be included using their absolute first drug administration date.
- Time in study during one or more treatment pauses will be summarized separately. This is equivalent to the white period of hiatus.

Figure 2: Exposure Calculations



If subjects pause and re-start more than once, the same rules will be applied. The most recent green period will be counted as efficacy exposure.

14.3. Dosing Compliance for the LTE Period

For the LTE period, a compliance percentage will be calculated for each subject. The calculation for dosing compliance in the LTE period will be (number of doses taken / number of expected doses during the LTE period. Compliance for subjects who re-start will only be calculated using the number of expected doses for the re-started study period (defined in Section 10.4.1).

14.4. Dosing Exposure for the LTE Period

Treatment exposure for each subject will be calculated using the same algorithm as shown in Dosing Exposure for the DB period. However, the date of last drug administration will be defined as the date of last dose of study drug plus 7 days after completion of treatment. The date of first drug administration will be the first date of pegzilarginase administration in the LTE period. The same algorithm for pauses due to COVID-19 in the DB period (Section 14.2) will be followed for the LTE period.

15. PROTOCOL DEVIATIONS

Protocol deviations and a subset of important protocol deviations (ie, those used in determining the PP set) will be categorized by a Clinical Research Physician. A list of all protocol deviations for all subjects for the 24-week data cutoff analysis will be defined prior to unblinding. The list of protocol deviations and important protocol deviations for all subjects for final analysis will be finalized prior to the final database lock. A review of these protocol deviations and important protocol deviations will also be performed in a blinded fashion to determine inclusion in PP set.

Protocol deviations and important protocol deviations will be summarized by treatment group and analysis period. Summary tables showing the number and percent of subjects with at least 1 protocol deviation and important protocol deviations by each category of protocol deviations and important protocol deviations will be provided. A listing of subjects with protocol deviations and important protocol deviations will be provided. Important protocol deviations will be identified via a flag in the listing.

All deviations related to COVID-19 are considered important protocol deviations, and COVID-19 related protocol deviations will be presented by category. A separate listing of subjects with COVID-19–related protocol deviations will be provided.

16. DERIVATION OF EFFICACY MEASURES

16.1. Plasma Arginine Concentrations

The derivation for baseline arginine is defined in Section 10.4.2.

The final follow-up arginine value for 24 weeks (endpoint arginine value) will be defined as the mean of the last 4 prior-to-dosing logged values obtained during the 24-week DB period that meet the following criteria:

- The sample date occurred after the scheduled date for the 20th dose, and
- The prior 2 doses were administered as planned (i.e., at approximately 7 days and 14 days prior to collection of the 24-week 168-hour arginine sample)
- Note:
  - If at least 1, but fewer than 4, values meet the criteria, then only those values will be included in the mean.
  - If none of the values meet the criteria, then the last single post-baseline arginine value will be included.
  - If no post-baseline arginine values were obtained, then the change from Baseline will be imputed as 0.

16.2. 2-Minute Walk Test

The 2MWT is a performance outcome measure of mobility, categorized as a neuromotor assessment in the protocol. Results for this test will be recorded in the electronic case report form (eCRF) as distance walked during 2 minutes in meters. A responder analysis will be conducted for this measure. A subject will be considered a responder if there is an improvement of ≥9% from Baseline. Otherwise, the subject will be considered a non-responder.

16.3. Gross Motor Function Measures

The Gross Motor Function Measure (GMFM) is a clinical measure designed to evaluate gross motor function by observing the subject's ability to initiate and complete certain movements. For this study, this measure is categorized as a neuromotor assessment in the protocol, and only Parts D (standing) and E (walking, running, and jumping) will be assessed.

16.3.1. Gross Motor Function Measure E

Part E consists of 24 questions that will be used to calculate a Total Part E score (the final score is an integer between 0 and 72). Change from Baseline at Week 24 will be calculated for each measure as the Week 24 total score – the baseline score. A responder analysis will be performed for this endpoint. A subject will be considered a responder for the GMFM-E if there is an improvement in scoring by 4.0 for GMFCS I, 2.8 for GMFCS II, and 1.8 for GMFCS III (see Table 6), and a non-responder otherwise.

16.3.2. Gross Motor Function Measure D

Part D consists of 13 questions that will be used to calculate a Total Part D score (the final score is an integer between 0 and 39); Change from Baseline at Week 24 will be calculated for each measure as the Week 24 total score – the baseline score. A responder analysis will be performed for this endpoint. A subject will be considered a responder for the GMFM-D if there is an improvement in scoring by 2.4 for GMFCS I, 3.3 for GMFCS II and 1.5 for GMFCS III (see Table 6), and a non-responder otherwise.

16.4. Functional Mobility Assessment

The Functional Mobility Assessment (FMA) objective will be assessed using 4 different endpoints: the FMS over 5 meters, 50 meters, and 500 meters (FMS 5, FMS 50, and FMS 500) and the GFAQ.

16.4.1. Functional Mobility Scale

The FMS is a 6-point scale from Level 1 (uses wheelchair, stroller, scooter, shopping cart, wagon, or is carried OR walks for exercise only with highly specialized/supportive walker OR does limited stepping with significant support/assistance from another person) to Level 6 (independent walking and running on all surfaces without assistive devices or help from another person) that assesses the need for assistive devices for walks of 3 different lengths: 5 meters, 50 meters, and 500 meters. Change from Baseline at Week 24 for each FMS (FMS 5, FMS 50, FMS 500) will be calculated as the Week 24 score – the baseline score. A responder analysis will be conducted. For this study, a subject will be considered a responder if there is at least a 1-level improvement from Baseline. If there is no improvement, or the subjects level decreases, the subject will be defined as a non-responder. Each distance (5, 50, and 500 meters) will be assessed separately.

16.4.2. Gillette Functional Assessment Questionnaire

The GFAQ is a parent/caregiver assessment consisting of a single question describing a child's ability to walk using a 10-point scale from Level 1 (cannot take any steps at all) to Level 10 (walks, runs, and climbs on uneven terrain and does stairs without difficulty or assistance; is typically able to keep up with peers). Change from Baseline at Week 24 for the GFAQ will be calculated as the Week 24 score – the baseline score. A responder analysis will be conducted. For this study, a subject will be considered a responder if there is at least a 2-level improvement from Baseline. If there is a 1-level improvement, no improvement, the subjects level decreases, or the value is missing, the subject will be defined as a non-responder.

16.5. Vineland Adaptive Behavior Scales-II

The VABS-II will be administered to the subject (as appropriate) and/or to the parent or caregiver. The VABS-II is a scale designed to measure adaptive behavior of individuals from birth to age 90 years. The VABS-II contains 4 domains: communication, daily living skills, socialization, and motor skills. The 4 domains are made up of 11 subdomains in which the scores are added to form the domain composite scores. The 4 domain composite scores then combine to form the Adaptive Behavior Composite for those individuals aged birth to 6 years 11 months.

Three domain composite scores (communication, daily living skills, and socialization) combine to form the Adaptive Behavior Composite for those ages 7 through 90 years.

For the purpose of this study, the 4 domains and the Adaptive Behavior Composite score will be assessed. Change from Baseline at Week 24 will be analyzed for each of the 4 domains and for the Adaptive Behavior Composite and defined as the Week 24 score-the baseline score. A responder analysis will be conducted. A subject will be defined as a responder if there is at least a 7.5-point improvement in the subject's standard score (range 20 – 140) as compared with Baseline. If there is a <7.5-point improvement, no improvement, the subject's score decreases, or the value is missing, the subject will be defined as a non-responder. Each of the 5 components will be assessed separately, regarding responder/non-responder status.

16.6. Composite Clinical Outcome: 2MWT, GMFM Part D, and GMFM Part E

The composite clinical tertiary outcome will be derived from the 2MWT, GMFM-D, and GMFM-E values at Week 24. The clinical response definitions are detailed below in Table 6.

Table 6: Clinical Response Definitions

Assessment	Definition of Component Improvement at Week 24 (Change from Baseline)	Definition of Component Worsening at Week 24 (Change from Baseline)
2MWT (distance walked in m)	Improvement by ≥9%	Worsening by ≥9%
GMFM-D	GMFCS I GMFCS II GMFCS III 2.4 3.3 1.5	GMFCS I GMFCS II GMFCS III -2.4 -3.3 -1.5
GMFM-E	GMFCS I GMFCS II GMFCS III 4.0 2.8 1.8	GMFCS I GMFCS II GMFCS III -4.0 -2.8 -1.8

Subjects with missing data for Week 24 in an assessment will be defined as neither improved nor worsened in that assessment.

To be classified as a responder for the composite clinical outcome, subjects must have an improvement in at least 1 component assessment with no worsening in any other components. An additional responder analysis will be conducted with a responder defined as having an improvement in at least 1 component assessment regardless of worsening in any other component. Additionally, the frequencies of subjects achieving response (using the criteria defined in Table 6 for improvement) for ≥1, ≥2, or 3 of the categories will be summarized.

In addition, a composite rank score will be created. For this measure, the change from Baseline will be measured for each of the 3 endpoints. Each of these values will then be separately rank ordered, and a summed rank score will then be determined for each subject (i.e., a subject rank score will be determined from summing 3 separate rank values). A Wilcoxon Rank Sum test will then be performed on the rank scores to compare the treatment groups.

Additional analyses may be performed based upon the outcome of separate clinical outcome assessment (COA) analysis and the determination of meaningful change thresholds (MCTs) used to define responders for these endpoints (see Section 18.13).

16.7. Modified Ashworth Scale

The MAS was developed to assess the spasticity of subjects with central nervous system lesions and is used to measure the resistance to passive movement about a joint due to spasticity. The muscle groups assessed include left and right of each of the following: biceps, triceps, hamstrings, and quadriceps. The scale utilizes a scoring scale of 0 (no spasticity) to 4 (total rigidity) with 6 scoring choices as shown in Table 7.

Table 7: Modified Ashworth Scale Definitions

Score	Modified Ashworth Scale
0	No increase in tone
1	Slight increase in tone manifested by a catch and release or by minimal resistance at the end of the ROM when moved in flexion or extension
1+	Slight increase in tone manifested by a catch, followed by a minimal resistance throughout the remainder (less than half) of the ROM
2	More marked increase in tone through most of the ROM, but affected parts easily moved
3	Considerable increase in tone; passive movement difficult
4	Limb rigid in flexion or extension

Abbreviations: ROM=range of motion.

Missing data will not be replaced or imputed for this measure.

16.8. Assessments of Caregiver and Clinician Global Impressions

16.8.1. Caregiver and Clinician Global Impressions of Change

The Caregiver and Clinician Global Impressions of Change are questions independently administered to the subject's caregiver and clinician regarding their impression of change in quality of mobility and in adaptive behavior during the DB portion of the study. Both the Caregiver and Clinician Global Impressions of Change include 8 questions which have the following response choices:

- Much worse
- Worse
- No clear change
- Better
- Much better
- Answer Not Selected

16.8.2. Caregiver and Clinician Global Impressions of Severity

The Caregiver and Clinician Global Impressions of Severity are single questions independently administered to the subject's caregiver and clinician regarding their impression of the current severity of the subject's deficits in mobility and adaptive behavior. Both the Caregiver and

Clinician Global Impressions of Severity include 8 questions that have the following response choices:

- Very much worse than others
- Much worse than others
- Worse than others
- A little bit worse than others
- No clear difference from others
- Answer Not Selected

16.9. 9-Hole Pegboard Test

Manual dexterity (hand function) is a fine motor skill that represents an individual's ability to coordinate the fingers and manipulate objects in a timely manner. The NIH Toolbox contains the 9-Hole Pegboard Dexterity Test, which is a simple test of manual dexterity for assessment of fine motor skills. The test records the time required for the participants to accurately place and remove 9 plastic pegs into a plastic pegboard.

Data will be collected and summarized descriptively across all age groups for both the dominant and non-dominant hand. Data will be reported as Z-scores for this measure. The Z-score is obtained by subtracting the mean of the reference population from the test result and then dividing by the SD of the reference population.

16.10. Quality-of-Life Assessments

The QoL assessments include the PedsQL, the SF-36, and the ZBI-12. Subjects (or parents) will complete either the PedsQL (2 to 18 years) or the SF-36 (19 years or greater) questionnaire. All caregivers will complete the ZBI-12.

16.10.1. Pediatric Quality of Life

The 23-item PedsQL Generic Core Scales were designed to measure the core dimensions of health as delineated by WHO. The 4 Multidimensional Scales are Physical Functioning (8 items), Emotional Functioning (5 items), Social Functioning (5 items), and School Functioning (5 items). The 3 Summary Scores are Total Scale Score (23 items), Physical Health Summary Score (8 items), and Psychosocial Health Summary Score (15 items).

The tool is developmentally appropriate for ages 2 to 18 years and can be administered in 7 different ways: Child Self-Report: ages 5 to 7 years, 8 to 12 years, and 13 to 18 years; Parent Proxy Report: ages 2 to 4 years, 5 to 7 years, 8 to 12 years, and 13 to 18 years. The investigator/ testing psychologist will determine whether a child is able to complete the self-report at study Baseline. For children where they are not able, or it is not appropriate for them to complete the self-report (e.g., children 2 to 4 years old), the Parent Proxy Report will be used.

Items are scored on a 5-point Likert scale from 0 (Never) to 4 (Almost always) with no weighting. Scores are then transformed on a scale from 0 to 100. Items are then reversed scored and linearly transformed to a 0-100 scale as follows: 0=100. 1=75, 2=50, 3=25, 4=0. Mean scores will be calculated over the number of items answered. A total score is calculated as the

sum of all items over the number of items answered on all scales. If >50% of items in a scale are missing, then the scale scores will not be computed.

This data will be summarized by dimension across all age groups.

16.10.2. 36-Item Short Form Health Survey

The SF-36 assesses 8 health concepts: limitations in physical activities because of health problems; limitations in social activities because of physical or emotional problems; limitations in usual role activities because of physical health problems; bodily pain; general mental health (psychological distress and well-being); limitations in usual role activities because of emotional problems; vitality (energy and fatigue); and general health perceptions. The SF-36 asks for participants to reply to questions according to how they have felt over the previous 4 weeks.

The scoring algorithms for the SF-36 individual scales and component summary scores are described in the User's Manual (Ware 1993). All scales and component summary measures are scored such that a higher score indicates better functioning. Norm-based scores will be used for the analysis of all scales and component summary measures.

Scoring the SF-36 is a 2-step process. First, scores are generated from 0 to 100, such that a high score defines a more favorable health status. Then, items in the same scale are averaged together to create the 8 scale scores. Missing data are not considered when calculating the scale scores. Hence, scale scores represent the average for all items in the scale that the subject answered.

The SF-36 will only be presented in a listing due to too few subjects having completed the assessment.

16.10.3. Zarit Burden Interview-12

The ZBI-12 is a 12-item, short version of the ZBI that is used to describe Caregiver burden. It consists of 12 items in 2 domains: personal strain and role strain. Each question is scored on a 5-point Likert scale from 0 to 4 (Never, Rarely, Sometimes, Quite frequently, Almost always), with higher scores representing higher feeling of burden. Thus, the overall score range is 0-48. Missing data are not considered when calculating the scores.

16.10.4. Neurocognitive Assessments

A subject who completes a baseline assessment will continue with that same assessment during follow-up, even if they are out of age range for the test during follow-up.

This study utilizes 4 separate assessments to assess neurocognitive functioning, each dependent on age. Below is a list of the 4 assessments with corresponding age ranges:

- Wechsler Adult Intelligence Scale IV (WAIS-IV): 16 years and older
- Wechsler Intelligence Scale for Children V (WISC-V): 6 to 16 years and 11 months
- Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV): 2.5 to 7.6 years
- Bayley Scales of Infant and Toddler Development III (BSID-III): 2 to 3.5 years

Each test has its own set of domain/composite scores appropriate for a specific age group. The 3 Wechsler tests all report a full-scale intelligence quotient (IQ) score.

The WAIS-IV test has 10 core subtests and 5 supplemental subtests. The test measures abilities in 4 main categories. These include reasoning, retention of information, processing and organization of information, and verbal comprehension. Factors such as creativity, individuality, or judgment are not incorporated into the test. Each category is scored individually, and the composite score is used to obtain the IQ score.

The WISC-V generates a Full-Scale IQ (formerly known as an IQ score) that represents a child's general intellectual ability. It also provides 5 primary index scores: Verbal Comprehension Index, Visual Spatial Index, Fluid Reasoning Index, Working Memory Index, and Processing Speed Index. These indices represent a person's abilities in discrete cognitive domains. Five ancillary composite scores can be derived from various combinations of primary or primary and secondary subtests.

The General Ability Index for this test is calculated from the General Ability Sum of Scaled Scores, which is the sum of scaled scores for 3 Verbal Comprehension subtests (i.e., Vocabulary, Comprehension, and Similarities) and 3 Perceptual Reasoning subtests (i.e., Block Design, Matrix Reasoning, and Picture Concepts).

The WPPSI is an intelligence test designed for children ages 2 years 6 months to 7 years 7 months. It provides subtest and composite scores that represent intellectual functioning in verbal and performance cognitive domains, as well as a composite score that represents a child's general intellectual ability (i.e., Full-Scale IQ). It consists of 14 subtests. They are designated as 1 of 3 types: core, supplemental, or optional. The core subtests are required for the computation of the verbal, performance, and Full-Scale IQ.

The BSID-III is a performance-based clinician-reported outcome assessment for use in children ages 2 to 3.5 years. Scales include Cognitive, Language (Receptive Communication and Expressive Communication subscales), and Motor (Gross and Fine Motor subscales). Each (sub)scale yields a total raw score which is then standardized according to the subject's chronological age (scaled scores).

Scoring

Number of items depends on which set is given in each domain. The number of questions administered depends in part on the child's age and on their developmental level.

The total raw scores for the Cognitive Scale and the Receptive Communication, Expressive Communication, Fine Motor, and Gross Motor subscales are the sums of the number of points earned for a (sub)scale.

The raw scores cannot be accurately compared to each other as each subscale has a different number of items resulting in a different range of possible scores. Therefore, the comparisons are best based on derived scores: age-based scaled scores or composite scores.

Scaled scores represent a child's performance on a (sub)scale relative to the same age peers. They are derived from the total raw scores on each of the subscale and are scaled to a metric with a range of 1 to 19, a mean of 10 and a SD of 3.

Composite scores are scaled to a metric with a mean of 100 and an SD of 15 and range from 40 to 160.

For the Language and Motor scale, composite scores are derived from the sums of age-corrected scaled scores. For each composite, the distribution of the sum of scaled scores is used to derive corresponding percentiles which are converted to composite scores with a mean of 100 and an SD of 15.

For the Cognitive subdomain, the scaled score to composite equivalent is a linear conversion from one scale (mean=10, SD=3) to another (mean=100, SD=15).

Missing data will not be imputed. Only a summary of the Combined Full-Scale IQ Score for subjects who received either the WAIS-IV, WISC-V, or WPPSI-IV will be provided in a table. In addition, a listing of scaled scores and composite scores will be provided.

16.11. Diet Management Plan and 3-Day Diary Diet Record

Subjects will be maintained on the dietary and ammonia scavenger regimen prescribed prior to enrollment by the treating investigator throughout the study. The prescribed regimen will not be modified during the study unless clinically indicated in the opinion of the investigator. The rationale for any change, and the date and nature of any significant (>±15% of Baseline) prescribed changes, will be recorded in the eCRF.

Subjects will be required to maintain dietary protein intake levels that are consistent with their baseline levels for the entire duration of the DB period and the 8-week blinded portion of the open-label LTE period. A consistent diet is defined as one in which the prescribed and consumed natural intact protein, medical food EAA, and calories are <±15% from Baseline. Any changes should also be recorded in the eCRF.

Subjects and/or their caregivers will be instructed to record all dietary intake by the study subject (natural food, medical food, and essential amino acids supplementation) for 3 consecutive days prior to all clinic visits during the screening period and before DB Visit IDs 6, 12, 18, and 24. At these visits, the prescribed regimen will also be recorded. The date and time of the last dose of essential amino acids must be recorded for each day of the 3-day diary diet record.

Compliance with the prescribed diet will be defined as the actual caloric intake being within 15% of what was prescribed at Baseline. Compliance will be determined for Total natural protein, Total EAA protein, and Total protein (EAA plus natural) separately. The frequencies of subjects whose protein is >15% higher than their prescribed diet and those who are >15% less than their prescribed diet at baseline will be summarized.

16.12. Growth Assessments

Growth assessments using Z-scores, including height (cm), height measurement (standing or supine), head circumference (pediatric subjects up to 18 years of age) (cm), and weight (kg) will be collected at visits shown on the Schedule of Assessments in the protocol. Growth centiles will be derived based on the collected data. The Z-score is obtained by subtracting the mean of the reference population from the test result and then dividing by the SD of the reference population.

16.13. Electroencephalogram

The electroencephalogram (EEG) is defined as electrical activity of an alternating type recorded from the scalp surface after being picked up by metal electrodes and conductive media. All

subjects will be consented for Baseline and follow-up routine (i.e., not sleep, ambulatory or video telemetry), outpatient EEG recordings. Screening and Week 24 recordings will be performed for all subjects and summarized as Normal, Abnormal Not Clinically Significant, Abnormal Clinically Significant.

16.14. Psychometric (Clinical Outcomes Assessment) Analyses

Blinded psychometric analyses to further inform the responder threshold determination for the key secondary and other secondary COA endpoints will be conducted by a separate vendor and summarized in a separate report. These include the 2MWT, GMFM Parts D and E, FMA, and VABS-II domains (Adaptive Behavior Composite, Communication Domain, Daily Living Skills Domain, Socialization Domain, and Motor Skills Domain). These MCID thresholds may be used to update the responder analyses.

17. PRIMARY AND SECONDARY ANALYSES

All efficacy analyses will be conducted using the FAS and summarized by randomized treatment. Sensitivity analyses for the primary and key secondary endpoints will be conducted using the PP set and summarized by randomized treatment as described in Section 17.1 and Section 17.2. Data for all efficacy analyses will be listed.

17.1. Primary Analysis and Sensitivity Analyses of the Primary Endpoint

The primary objective is to evaluate the efficacy of pegzilarginase versus placebo based on plasma arginine concentrations after 24 weeks of study treatment compared to Baseline.

The derivation for baseline arginine is defined in Section 10.4.2.

The final follow-up arginine value for 24 weeks (endpoint arginine value) will be defined as the mean of the last 4 prior-to-dosing (values collected at 168 hours post-dose) logged values obtained during the DB period that meet the criteria outlined in Section 16.1.

The primary analysis of the primary endpoint is change from Baseline comparing the baseline logged arginine value to the endpoint logged arginine value using an ANCOVA model. Results will be presented as geometric mean values (from the first model below analyzing the actual values at the timepoint), ratios to Baseline values, and changes with 95% CIs (from the second model below). The treatment effect will be presented as a relative ratio to Baseline and change with 95% CIs and a 2-sided p-value (from either model which will both give the same treatment effect because of the inclusion of BASE as a covariate). The baseline arginine value will be included as a covariate in the ANCOVA model.

The following SAS code, using the SAS MIXED procedure, or equivalent is representative of a model to be used on the logged data. LS Mean estimates and 95% CIs will be exponentiated before being presented:

```
proc mixed method=reml;
class TREATMENT VISIT USUBJID;
model AVAL = BASE TREATMENT VISIT TREATMENT*VISIT / ddfm=kr;
repeated VISIT / type=UN subject=USUBJID;
lsmeans TREATMENT*VISIT / om diff cl;
run;


proc mixed method=reml;
class TREATMENT VISIT USUBJID;
model CHG = BASE TREATMENT VISIT TREATMENT*VISIT / ddfm=kr;
repeated VISIT / type=UN subject=USUBJID;
lsmeans TREATMENT*VISIT / om diff cl;
```

As a sensitivity analysis, change from Baseline to final follow-up arginine level will be compared between pegzilarginase and placebo using a Wilcoxon rank sum test.

In order to provide a sensitivity analysis, the primary analysis (ie., change from baseline comparing the baseline actual arginine value to the endpoint actual arginine value using an ANCOVA model) will be performed using the raw data.

The endpoint value (logged measurements, averaged and exponentiated [i.e., a geometric mean]), change from Baseline and ratio to Baseline will be summarized by randomized treatment.

Arginine values, change from Baseline, and ratio to Baseline will be summarized by randomized treatment and by visit.

The primary analysis, sensitivity analyses, and summaries will be performed on the FAS and will be repeated on the PP set.

The analysis and summaries will be repeated by subgroup on the subgroups defined in Section 10.6 on the FAS only, assuming there are sufficient subjects in the subgroups.

17.2. Key Secondary Analyses and Sensitivity Analyses of the Key Secondary Endpoints

There are 2 key secondary endpoints: mean change from Baseline at Week 24 in 2MWT and mean change from Baseline at Week 24 in GMFM-E. Both endpoints will be analyzed using the same method and summarized in the same way.

The endpoint will be analyzed using an MMRM using data from Week 12 and Week 24, LS mean estimates, and differences between treatments from both time points will be presented, but the test of difference from 24 weeks is the key secondary endpoint analysis.

Results will be presented as both actual value and change from Baseline with 95% CIs (97.5% if only one of the 2 key secondary endpoints is significant from the Hochberg procedure). The treatment effect will be presented as a difference in change from Baseline with a 95% CI (97.5% if only one of the 2 key secondary endpoints is significant from the Hochberg procedure) and a 2-sided p-value. The MMRM will include treatment, visit, treatment-by-visit interaction, and the baseline value as a covariate.

The following SAS code, used for the MIXED procedure, or equivalent may be used for modelling the actual value and the change from Baseline value:

```
proc mixed method=reml;
class TREATMENT VISIT USUBJID;
model AVAL = BASE TREATMENT VISIT TREATMENT*VISIT / ddfm=kr;
repeated VISIT / type=UN subject=USUBJID;
lsmeans TREATMENT*VISIT / om diff cl;
run;


proc mixed method=reml;
```

```
class TREATMENT VISIT USUBJID;
model CHG = BASE TREATMENT VISIT TREATMENT*VISIT / ddfm=kr;
repeated VISIT / type=UN subject=USUBJID;
lsmeans TREATMENT*VISIT / om diff cl;
run;
```

If there are not enough data for an unstructured covariance matrix to be used, then a compound symmetry covariance matrix will be used instead (type=CS).

Endpoint values and change from Baseline will be summarized by randomized treatment and by visit.

The key secondary analyses and summaries will be performed on the FAS. The analyses and summaries will be repeated on the PP set.

The key secondary analyses and summaries will be repeated by subgroups, as defined in Section 10.6 on the FAS only.

The MMRM produces estimates which are consistent with a MAR assumption for missing data. To assess the reliability of results arising with the MMRM, an additional tipping point sensitivity analysis using multiple imputation (MI) methods will be performed. The tipping point analysis will assess the potential impact of informative missingness by progressively penalizing subjects with missing data in the selected treatment group. The goal is to find the level of penalization that results in loss of statistical significance on the key secondary endpoint. The more extreme the penalization required to render the key secondary endpoint p-value non-significant, the greater one can consider the reliability of the principal MMRM analysis. The methodology will be similar to that provided by Ouyang et al (2017), using the MI and MIANALYZE procedures in SAS.

In order to perform a sensitivity analysis, the key secondary analyses (ie, mean change from Baseline at Week 24 in 2MWT and mean change from Baseline at Week 24 in GMFM-E) will be tested using a Wilcoxon rank sum test.

17.3. Other Secondary Analyses

All other secondary analyses will be performed on the FAS only.

For all responder analyses, if data is missing at the timepoint of interest, the subject will be a non-responder for the purpose of the analysis.

17.3.1. Compare Treatment Groups Whose Endpoint Arginine Value Falls Below 200 µM

Using the endpoint arginine value calculated for the primary objective (Section 17.1), subjects will be identified as a responder if their endpoint arginine value is <200 µM. If a subject's endpoint arginine value is ≥200 µM, the subject will be a non-responder.

A 2-sided Fisher's exact test will be used to compare clinical responders versus non-responders by treatment group at Week 24, and responders and non-responders for this definition will be tabulated.

17.3.2. Compare Treatment Groups Whose Endpoint Arginine Value Falls Within 40 to 115 µM

Using the endpoint arginine value calculated for the primary objective (Section 17.1), subjects will be identified as a responder if their endpoint arginine value is ≥40 µM and ≤115 µM. If a subject's endpoint arginine value is outside this range, the subject will be a non-responder.

A 2-sided Fisher's exact test will be used to compare clinical responders versus non-responders by treatment group at 24 weeks, and responders and non-responders for this definition will be tabulated.

17.3.3. Compare Pegzilarginase With Placebo for Changes in GCs and Ornithine

Change from Baseline in GCs and ornithine after 24 weeks of study treatment will be analyzed and summarized on logged data using the same methods as the primary endpoint described in Section 17.1.

17.3.4. Compare Pegzilarginase With Placebo With Respect to Other Aspects of Mobility

Mean change from Baseline at Week 24 in the GMFM-D will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in FMS 5 will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in FMS 50 will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in FMS 500 will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in the GFAQ will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

17.3.5. Compare Pegzilarginase With Placebo With Respect to Adaptive Behavior

Mean change from Baseline at Week 24 in the VABS-II Adaptive Behavior Composite will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in the VABS-II Communication Domain will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in the VABS-II Daily Living Skills Domain will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in the VABS-II Socialization Domain will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

Mean change from Baseline at Week 24 in the VABS-II Motor Skills Domain will be analyzed and summarized using the same methods as the key secondary endpoints as described in Section 17.2.

17.3.6. Evaluate the Safety and Immunogenicity of Pegzilarginase

Safety reports, more specifically AEs, are detailed in Section 19.1. Additional ADA analyses will be included as part of the PK/PD plan, which will be included in a separate report.

17.3.7. Characterize the Pharmacokinetic Profile of Pegzilarginase

A separate PK/PD analysis plan will be developed, and results will be presented in a separate report.

18. TERTIARY ANALYSES

All tertiary analyses will be performed on the FAS only.

For all responder analyses, if data is missing at the timepoint of interest, the subject will be a non-responder for the purpose of the analysis.

18.1. Compare the Response on Pegzilarginase With Placebo With Respect to Mobility Assessments (2MWT, GMFM-E, and GMFM-D)

A 2-sided Fisher's exact test will be used to assess mobility (2MWT, GMFM-D, and GMFM-E) responders versus non-responders as described in Section 16.6 by treatment group at Week 24. To be classified as a responder for the composite clinical outcome, subjects must have an improvement in at least 1 component assessment with no worsening in any other components. An additional responder analysis will be conducted with a responder defined as having an improvement in at least 1 component assessment regardless of worsening in any other component. Additionally, the frequencies of subjects achieving response (using the criteria defined in Table 6 for improvement) for ≥1, ≥2, or 3 of the categories will be summarized.

Additionally, the component classifications (improvement, worsening, and neither improved nor worsened) will be tabulated.

18.2. Compare the Response on Pegzilarginase With Placebo With Respect to the FMS 5, FMS 50, FMS 500, and GFAQ

Response for the FMS and GFAQ is defined as a 1-level change from Baseline at any distance and a 2-level change from Baseline at any distance, respectively.

Week 24 responders for FMS 5, FMS 50, and FMS 500 will be compared. A 2-sided Fisher's exact test will be used to compare by treatment group.

A 2-sided Fisher's exact test will be used to compare GFAQ responders versus non-responders by treatment group at Week 24.

18.3. Compare the Response on Pegzilarginase With Placebo With Respect to the VABS-II Overall Composite and Individual Domains

Response for VABS-II is defined as an improvement by ≥7.5 points.

Week 24 responders for the following will be compared:

- A 2-sided Fisher's exact test will be used to compare Adaptive Behavior Composite responders versus non-responders by treatment group at Week 24, and responders and non-responders for this definition will be tabulated for Week 24.
- A 2-sided Fisher's exact test will be used to compare Communication Domain responders versus non-responders by treatment group at Week 24, and responders and non-responders for this definition will be tabulated for Week 24.
- A 2-sided Fisher's exact test will be used to compare Daily Living Skills Domain responders versus non-responders by treatment group at Week 24, and responders and non-responders for this definition will be tabulated for Week 24.

- A 2-sided Fisher's exact test will be used to compare Socialization Domain responders versus non-responders by treatment group at Week 24, and responders and non-responders for this definition will be tabulated for Week 24.
- A 2-sided Fisher's exact test will be used to compare Motor Skills Domain responders versus non-responders by treatment group at Week 24, and responders and non-responders for this definition will be tabulated for Week 24.

18.4. Compare Pegzilarginase With Placebo With Respect to Improvement of Spasticity and Objective Measures of Neurological/Neuromotor Manifestations

Change from Baseline for functional and spasticity measurements using the MAS (see Section 16.7) will be summarized by treatment group, visit, and upper (left and right biceps, left and right triceps) and lower muscle groups (left and right hamstrings, left and right quadriceps) using descriptive statistics. Subjects may be excluded from the analysis if Botulinum toxin was used during the study.

Muscle group scores of 1+ will be converted to 1.5 for computational purposes. Data points collected from subjects who initiate concomitant medication for spasticity (e.g., Botox®) may be excluded. To examine lower body spasticity, the sum of the MAS scores for the lower body muscle groups (right hamstrings, left hamstrings, left quadriceps, right quadriceps) will be calculated. A sum score per muscle group will be calculated as the sum of MAS score for lower body muscle groups divided by the number of impacted lower body muscle groups at baseline. These scores and changes from Baseline will be presented by visit. Improvement in spasticity corresponds to a reduction in the MAS score per muscle group from Baseline based on the 4 lower body muscle groups at the respective week. Worsening in spasticity corresponds to an increase in the MAS score per muscle group from baseline scores based on the 4 lower body muscle groups at the respective visit.

The same analyses will be performed separately for the upper muscle groups (left and right biceps, left and right triceps).

18.5. Describe the Impact of Global Impression of Change/Severity on Motor Function and Adaptive Behavior

The Caregiver and Clinician Global Impressions of Change and Global Impression of Severity are questions independently administered to the subject's caregiver and clinician regarding their impression of change in quality of mobility and in adaptive behavior, and their impression of severity of the subject's deficits in mobility and adaptive behavior, respectively. These questions are asked at each post-baseline visit in the DB period during which each of the following 4 assessments are administered: 2MWT, GMFM-D, GMFM-E, and/or VABS-II. Of the 3 questions asked, the first 2 are for both caregivers and clinicians, while the third question, which has 6 parts, is for caregivers only.

The questions for Impression of Change and for Impression of Severity are similar in nature, and the responses are also similar. The 5 possible responses for the Impression of Change questions range from Much Worse to Much Better, while the 5 responses for the Impression of Severity

questions range from Very Much Worse Than Others to No Clear Difference From Others. Summaries of these responses will be reported together due to an overlap in questions.

Both the Caregiver and Clinician Global Impressions of Change and Caregiver and Clinician Global Impressions of Severity (Section 16.6) will be summarized independently for each of the 4 sets of questionnaires. Descriptive statistics will be reported for each response by treatment group and visit.

18.6. Describe the Impact of Pegzilarginase on Fine Motor Behavior

The 9-Hole Pegboard test will be used to assess this objective (Section 16.9). Descriptive statistics summarizing the time required for subjects to accurately place and remove 9 plastic pegs into a plastic pegboard will be reported independently for both the dominant and non-dominant hand. Results will be presented using Z-score data, by treatment group. The Z-score is obtained by subtracting the mean of the reference population from the test result and then dividing by the SD of the reference population.

18.7. Describe the Impact of Pegzilarginase on QoL

All QoL assessments (Section 16.10) will be summarized using descriptive statistics by treatment group. Below is a list of possible assessments, with evaluable parameters, for this objective:

- PedsQL (2 to 18 years) – Summary statistics will be reported on the transformed scores for each of the following parameters: Physical, Emotional, Social, and School. Summarize the summary scores for Psychosocial Health and Physical Health. Below is a list of all PedsQL versions:
  - PedsQL (toddler 2 to 4 years) – Parent report
  - PedsQL (child 5 to 7 years)
  - PedsQL (child 8 to 12 years)
  - PedsQL (teen 13 to 18 years)
  - PedsQL (parent 5 to 7 years)
  - PedsQL (parent 8 to 12 years)
  - PedsQL (parent 13 to 18 years)

  This data will be summarized by treatment group and dimension across all age groups for the placebo-controlled DB period. Additionally, subject listings will be provided.
- SF-36 (for subjects aged 19 years or greater):
  - Physical Functioning
  - Role Functioning/Physical
  - Role Functioning/Emotional
  - Energy/Fatigue

  - Emotional Well-being
  - Social Functioning
  - Pain
  - General Health

  Scoring the SF-36 is a 2-step process. First, scores are generated from 0 to 100 such that a high score defines a more favorable health status. Then items in the same scale are averaged together to create the 8 scale scores. Missing data are not considered when calculating the scale scores. Hence, scale scores represent the average for all items in the scale that the subject answered.

  Due to the small number of subjects (i.e., < 4 subjects) completing the SF-36, no summary table will be presented. A subject listing will be provided.

- ZBI-12 (all caregivers) – Summary statistics will be reported on the Total ZBI-12 Score for the DB period. Additionally, a subject listing will be provided.

18.8. Describe the Impact of Pegzilarginase on Neurocognition

All neurocognitive assessments (Section 16.11) will be summarized by treatment group during the placebo-controlled DB period using descriptive statistics. Below is a list of possible assessments, with evaluable parameters, for this objective:

- WAIS-IV – Summary statistics will be presented on the following 5 composite scores: Verbal Comprehension, Perceptive Reasoning, Working Memory, Processing Speed, and Full-Scale IQ.
- WISC-V – Summary statistics will be presented on the following 6 composite scores: Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, Processing Speed, and Full-Scale IQ.
- WPPSI-IV – Summary statistics will be presented on the following 6 (4, if <4 years of age) composite scores: Verbal Comprehension Index, Visual Spatial Index, Fluid Reasoning Index (not available if <4 years of age), Working Memory Index, Processing Speed Index (not available if <4 years of age), and Full-Scale IQ.
- BSID-III – Summary statistics will be presented on the following 3 composite scores: Cognitive, Language, and Motor.
  - Note: If there are <5 subjects who receive the BSID-III, then only results will be listed.

Only a combined summary report of Full-Scale IQ will be presented for those subjects who receive either the WAIS-IV, WISC-V, or WPPSI-IV. The IQ score can range 40 to 160, with a score of 90 to 109 considered ‘average’. Subject listings will be provided for each of the assessments.

18.9. Describe the Diets Maintained by Subjects

Descriptive statistics will be reported for the placebo-controlled DB period on a subject's 3-day diet record (Section 16.12) as captured on the eCRF. The following 6 parameters will be summarized by visit and by treatment group:

- Total prescribed natural protein/day (grams)
- Total consumed natural protein/day (grams)
- Total prescribed EAA (protein)/day (grams)
- Total consumed EAA (protein)/day (grams)
- Total prescribed protein/ day (EAA plus natural) (grams)
- Total consumed protein/ day (EAA plus natural) (grams)

The proportion of compliant and non-compliant (ie, those ≥15% below the prescribed levels and those ≥15% above the prescribed levels at baseline) subjects for each category will be summarized by treatment group and visit for the DB period. Subject listings, including compliance, will also be provided.

18.10. Describe the Effect of Pegzilarginase Associated With Plasma Ammonia Levels

Plasma ammonia will be summarized during the placebo-controlled DB and LTE periods using descriptive statistics by visit and by treatment group. The data reported will be the ammonia result in micromoles (μM).

18.11. Evaluate the Effects of Pegzilarginase on Growth in Pediatric Subjects

Growth assessment data will be collected (Section 16.13) and Z-scores for height, weight, and head circumference will be computed using the CDC growth curves for subjects 18 years of age or younger (CDC 2002). Descriptive statistics, including percentiles, will be reported during the placebo-controlled DB period by visit, and by treatment group. For subjects >18 years of age, age-normalized percentiles for height will be computed at Baseline only, and weight percentiles will not be computed.

18.12. Electroencephalogram

An EEG is used to evaluate electrical activity in the brain. All subjects will be consented for Baseline and follow-up routine, outpatient EEG recordings. Screening and Week 24 recordings will be performed for all subjects. An EEG result will be interpreted as normal, abnormal not clinically significant, or abnormal clinically significant. If a result is reported as abnormal clinically significant, then the abnormality will be reported.

A subject listing for abnormalities with a reported result of abnormal clinically significant will be listed for the DB analysis period. No other summary reports or listings will be generated for this data.

18.13. Psychometric (Clinical Outcome Assessment) Analyses

Blinded psychometric analyses to further inform the responder threshold determination for the key secondary and other secondary COA endpoints will be conducted by a separate vendor. These include the 2MWT, GMFM Parts D and E, FMA, and VABS-II domains (Adaptive Behavior Composite, Communication Domain, Daily Living Skills Domain, Socialization Domain, and Motor Skills Domain).

Anchor-based methods using the individual Caregiver and Clinician Global Assessments of Severity (static anchor) and Change (retrospective) as anchors, supplemented with both cumulative distribution function and probability density function curves, will be considered as the primary methods to derive the MCTs that will be used to define responders for each of the above endpoints. Results from the caregiver exit interviews will provide evidence supporting the selection of the MCTs. All COA-related endpoint analyses will be conducted on the DB period data. All psychometricians and analytic programmers will be blinded to clinical study results throughout this psychometric evaluation.

A separate psychometric analysis plan will be created to describe the psychometric analyses in more detail.

18.14. Characterize the Long-Term Treatment Effects of Pegzilarginase Administration Once Weekly

As previously described, the placebo-controlled DB analysis will be summarized by treatment. Since some additional data beyond the 24-week data cut-off date will be available at the time of the 24-week datacut, select safety and efficacy analyses will be summarized by treatment regimen through the LTE period on available subject data. Available data will also be included in listings.

18.15. Patient Profiles

To help describe subject journeys through the study, by-subject figures will be produced showing important subject information such as arginine values and dosing information, as well as the efficacy assessments of the 2MWT, GMFM-E, GMFM-D, MAS Lower Body, and FMA.

19. SAFETY ANALYSES

All safety analyses will be conducted using the FAS and summarized by actual treatment received (e.g., if the subject received at least 1dose of pegzilarginase, then the subject will be summarized in the pegzilarginase treatment group).

Safety will be assessed via descriptive statistics and point estimates. Unless otherwise specified, all safety analyses described in this section will be performed for both the DB and LTE periods. The safety analyses for the DB period will be presented by treatment group (placebo or pegzilarginase), while the analyses for the LTE period will be presented by treatment regimen (placebo - pegzilarginase (or Pbo-Peg) or pegzilarginase – pegzilarginase (or Peg-Peg).

For the analyses of AEs and marked abnormalities, the following point estimates are distinguished:

- Subject incidence
  - Subject incidence (i.e., percentage [%] used in a frequency summary) is defined as the number of subjects with the specific event divided by the number of subjects included in the analysis. Subjects with multiple occurrences of the specific event in the specific analysis period will be counted only once in the numerator.
- Exposure-adjusted incidence rate (EAIR) in subject-weeks
  - The EAIR in subject-weeks is defined as the number of subjects with the specific event divided by the total exposure time (in subject-weeks) among subjects included in the analysis. Subjects with multiple occurrences of the specific event in the specific analysis period will be counted only once in the numerator. The exposure time for a subject without the specific event is the treatment duration, whereas the exposure time for a subject with the specific event is the treatment duration up to the start date (inclusive) of the first occurrence of the specific event. The total exposure time in subject-weeks is calculated by dividing the sum of exposure time in days over all subjects included in the analysis by 7. The EAIR in subject-weeks is interpreted as the expected number of subjects with at least one occurrence of the specific event per subject-weeks of exposure to the study drug.

AEs and marked abnormalities will be summarized by subject incidence and EAIR for the DB/LTE combined period.

Descriptive statistics will be provided for ammonia and LFT laboratory values (continuous measurements) by treatment and visit, including the end of treatment visits. The baseline value, value at the timepoint, and change from Baseline will be summarized for subjects who have values at Baseline and at the timepoint. Data for vital signs, urinalysis, chemistry, coagulation and hematology labs will be provided in listings.

19.1. Adverse Events

An AE is defined as any untoward medical occurrence associated with the use of a drug in humans, whether the event is considered drug related. An AE may include intercurrent illnesses or injuries that represent an exacerbation of pre-existing conditions.

The reporting period for nonserious and serious adverse events (SAEs) is the period from the signing of informed consent continuing through the end of study visit. All AEs will be followed to resolution or until they become insignificant to further follow-up. Resolution is defined as the return to Baseline status or stabilization of the condition, with the expectation that it will remain chronic.

Treatment-emergent adverse events (TEAEs) are defined as any event that occurs or worsens on or after the day that randomized study treatment is initiated. Adverse events recorded on the eCRF that began prior to treatment will not be included in the summary tables but will be included in the AE data listings. All AEs will be coded using MedDRA Version 24.0 terminology or higher. Unless otherwise specified, AEs will be summarized by SOC, and PT, with SOCs presented in the order of overall subject incidence and then in alphabetical order and PTs within SOCs presented in descending order of overall subject incidence.

The severity of each AE will be assessed by the investigator or qualified designee using the categories defined as follows:

- **Mild –** Event usually transient, requires minimal or no treatment and does not generally interfere with the participant's daily activities.
- **Moderate –** Event usually alleviated with additional specific therapeutic intervention. The event interferes with usual activities of daily living, causing discomfort, but poses no significant or permanent risk of harm to the research participant result in a low level of inconvenience or concern with the therapeutic measures. Moderate events may cause some interference with functioning but poses no significant or permanent risk of harm to the participant.
- **Severe –** Event interferes a participant's usual daily activity and may require intensive therapeutic intervention. Of note, the term "severe" does not necessarily equate to "serious."

The relationship of each AE to the study drug, as described in the protocol, will be determined by the investigator using the following 5 categories:

- Not Related
- Unlikely Related
- Possibly Related
- Probably Related
- Definitely Related

Any AE that is classified as definitely related, probably related, or possibly related to study drug will be considered treatment-related. AEs will be considered unrelated to treatment by study drug if classified as not related or unlikely related. In the case of multiple occurrences of the same AE

reported by the same subject, each subject will be counted only once for each PT. In tables where the severity or relationship of the event to study drug is tabulated, the AE with the greatest severity or strongest relationship to the study drug will be the event counted. An AE with a missing severity and/or relationship to the medication will be classified as "Severe" and/or "Related," respectively, for the purpose of analysis and summarization.

An overall summary of AEs will be provided for the placebo-controlled DB and DB/LTE periods. The overall summary will include the incidence of TEAEs. The number of events and the numbers and percentages of subjects with any TEAE will be summarized by severity of the TEAEs, treatment-related TEAEs, TEAEs requiring dose reduction, TEAEs leading to dose interruption, TEAEs leading to discontinuation of study drug, SAEs, drug-related SAEs, and any TEAEs with a fatal outcome.

Overall, TEAEs will be summarized for the DB period with subject incidence in descending frequency as follows:

- Overall Summary of TEAEs
- TEAEs by SOC and PT Including Counts of Events and Frequency of Occurrence
- TEAEs by SOC and PT
- TEAEs by SOC, PT, and Maximum Severity
- Treatment-Related TEAEs by SOC, PT, and Maximum Severity
- TEAEs Occurring in ≥15% of Subjects by PT
- TEAEs Occurring in ≥ 15% of Subjects relative to Dose and PT
- Treatment-Related TEAEs Occurring in ≥15% Subjects by PT
- Treatment-Emergent SAEs by SOC and PT
- Treatment-Emergent SAEs by SOC, PT, and Maximum Severity
- Treatment-Emergent Treatment-Related SAEs by SOC and PT
- Hypersensitivity TEAEs by SOC and PT
- Injection Site Reaction TEAEs by SOC and PT
- Liver Function Test TEAEs by SOC and PT
- Hyperammonemia TEAEs by PT
- TEAEs Leading to Discontinuation of Study Drug by SOC and PT
- TEAEs Leading to Dose Interruption of Study Drug by SOC and PT
- TEAEs Leading to Dose Reduction
- SAEs Leading to Death by SOC and PT

In addition, TEAEs will be summarized for the LTE period with subject incidence in descending frequency as follows:

- Overall Summary of TEAEs

- TEAEs by SOC and PT Including Frequency of Occurrence
- Treatment-Related TEAEs by SOC and PT
- Treatment-Emergent SAEs by SOC and PT
- Treatment-Emergent Treatment-Related SAEs by SOC and PT
- Hypersensitivity TEAEs by SOC and PT
- Injection Site Reaction TEAEs by SOC and PT
- Liver Function Test TEAEs by SOC and PT
- Hyperammonemia TEAEs by PT

A listing will be presented for all TEAEs with details including, but not limited to, date of onset, severity, drug relationship, action taken, outcome, and route of administration of pegzilarginase at the time of AE onset.

19.1.1. Adverse Events of Special Interest

Hypersensitivity reactions, injection site reactions (ISRs), and hyperammonemic episodes are defined as AESIs for this clinical study. Separate summaries as line listings of AESIs, will be presented by PT and will include but not be limited to the following information: total number of events reported, seriousness, investigator causality, time to onset, severity, and subject numbers and event frequencies.

- **Hypersensitivity reactions:** are known effects of enzyme replacement therapy. The number of subjects reporting at least 1 hypersensitivity event along with the incidence of the individual PTs (as well as those events considered by the investigator as an HSR) will be summarized. A by-subject listing of hypersensitivity AEs will be presented. Additional analysis will be conducted of the narrow standardized MedDRA query (SMQ) for Hypersensitivity Reactions.
- **ISRs:** are known effects of enzyme replacement therapy given via SC injection. They are a local phenomenon defined as a constellation of symptoms, including swelling, erythema, pruritus, and pain around the site of the injection. The number of subjects reporting at least 1 ISR along with the incidence of the individual PTs will be summarized. A by-subject listing of ISR AEs will be presented. A separate table will be provided for ISR TEAEs, and a listing will be provided by subject. Additional analysis will be conducted of the MedDRA HLT Injection site reactions.
- **Hyperammonemia episodes:** Hyperammonemia is a known disease manifestation of ARG1-D. Because of the potential impact of hyperammonemic episodes on subject safety and the validity of clinical assessments, hyperammonemia is considered an AESI to enable tracking and more careful monitoring of contributing factors and the frequency/severity of these episodes. HA will be summarized as an Aeglea-defined ad-hoc query, which will include the MedDRA PTs of Hyperammonaemia, Hyperammonaemic crisis, and Hyperammonaemic encephalopathy. Ammonia increased was evaluated for the SMQ but was not included as investigators were

instructed to report any events of Ammonia increased that met the protocol specified definition as Hyperammonemia while the remaining PTs did not meet the definition. The number of subjects reporting at least 1 hyperammonemia TEAE and the incidence of the individual PTs will be summarized.

AESIs will be summarized by treatment group in the placebo-controlled DB period and by treatment regimen in the LTE period.

19.2. Physical and Neurological Examinations

The primary means of assessing physical examination results is through assessing AEs related to the physical examination.

No summaries of physical examination are planned, but a listing of abnormal changes in physical examination will be produced.

19.3. Vital Signs

Vital signs (including systolic and diastolic blood pressure, pulse, respiratory rate, and body temperature) will not be summarized in a table, but will presented in a by-subject listing

19.4. Electrocardiogram

A 12-lead ECG will be obtained after approximately 5 minutes of rest in the supine position using equipment at the site at the visits shown on the Schedule of Assessments in the protocol. Heart rate, PR, RR, QRS, QT, and QTc interval and formula (Bazett or Fridericia or other) will be collected. Both Bazett and Fridericia corrected QTc will be presented for each subject. Where available the value presented will be from the database; missing values will be calculated.

QTc intervals will be calculated using the following equations:

$$\text{Bazett: } QTc = QT / \sqrt{RR}$$
$$\text{Fridericia: } QTc = QT / RR^{1/3}$$

ECGs will be interpreted as normal, abnormal not clinically significant, or abnormal clinically significant. If a result is reported as abnormal clinically significant, then the abnormality will be reported. The number and percentage of subjects with QTc prolongation at any post-baseline visit will also be summarized:

- QTc >450 msec
- QTc >480 msec
- QTc >500 msec

A listing of ECG results as well as a listing including descriptions of ECG abnormalities will be produced. Investigator assessment of clinical significance will also be provided in the listing.

19.5. Clinical Laboratory Data

Clinical laboratory data for safety and efficacy will be obtained and forwarded to the central laboratory or processed locally according to the Schedule of Assessments in the protocol. Data

collected at specified visits will include the following: hematology, serum chemistry, other biomarkers (e.g., growth hormone, insulin-like growth factor 1), arginine, ornithine, GCs, amino acids, ammonia (local laboratory), urinalysis, coagulation, mutation analysis and arginase activity in RBC, phenylbutyrate metabolites, serum tryptase, and serum complement C3. Additional detail on each parameter is included in the protocol.

Continuous and categorical laboratory data will be summarized as appropriate. For quantitative measures, descriptive statistics will be used to summarize laboratory measurements based on the actual value, the maximum, the minimum, and the change from baseline value at each timepoint of assessment. A listing of serum tryptase and complement C3 will be produced for subjects who experienced a hypersensitivity reaction when available. All data will be presented in listings.

19.5.1. Ammonia Increases

Increased ammonia levels are defined as elevations in ammonia beyond the normal range that did not meet the prespecified protocol definition of Hyperammonemic episodes:

- ≥100 μM and
- 1 or more symptoms related to hyperammonemia and
- requiring hospitalization or emergency room management, with or without admission to the hospital

The methods for measuring plasma ammonia levels vary among individual local laboratories, and values obtained using different assay methods may not be interchangeable because the normal ranges and therapeutic target levels are dependent on the assay method used by the individual local laboratory. Therefore, normalized ammonia (μM) will be calculated using the formula:

Normalized ammonia (μM) = ammonia in μM x (35/ULN of the specified laboratory reference range for each assay).

Summaries of the event count and frequency of increases in ammonia levels will be produced for subjects with ammonia values of the upper limit of normal (ULN) to ≤100 μM, >100 μM to <250 μM, ≥250 μM to <500 μM, and ≥500 μM.

19.5.2. Liver Function Test Abnormalities

The incidence of the following liver function abnormality categories will be summarized for the liver function parameters of alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline Phosphatase (ALP), and Total Bilirubin. Hepatic Disorder TEAEs were evaluated and are defined using the MedDRA Standardized MedDRA Query (SMQ) narrow search for ‘Hepatic Disorder’.

These summaries based on post-baseline values will include the number and percentage of subjects in the categories listed in Table 8.

Table 8: Liver Function Test Abnormality Categories

ALT, AST or either ALT or AST
> 3 × ULN
> 5 × ULN
> 10 × ULN
> 20 × ULN
Total Bilirubin
> 2 × ULN
Combined LFTs
ALT or AST >3 × ULN, total bilirubin >2 × ULN and no ALP >2 × ULN*

Abbreviations: ALP=alkaline phosphatase; ALT=alanine aminotransferase; AST=aspartate aminotransferase; ULN=upper limit of normal.

For the assessment of liver function test abnormalities, all scheduled and unscheduled visits will be considered.

A listing of subjects who meet these LFT abnormality criteria will be presented.

20. PHARMACOKINETIC, PHARMACODYNAMIC, AND IMMUNOGENICITY ANALYSES

PK and PD analyses, as well as the immunogenicity assessment for this study, will be executed by independent vendors. The analysis plan(s) will be described in separate documents. A summary of the results will be presented in the main study CSR while additional details will be available in separate reports.

21. INTERIM ANALYSES

No interim analyses are planned during the DB portion of the study.

When all subjects have completed the DB portion of the study to LTE01 and the database to this timepoint has been frozen, the data up to that point will be formally unblinded and analyzed. Note that by the time the last evaluable subject has completed the placebo-controlled DB portion of the study and the Week 24 datacut is taken, many subjects will have data available beyond 24 weeks. Select efficacy analyses will be performed for all available data through 48 weeks. For those analyses, the pbo-peg treatment regimen will be compared against the peg-peg treatment regimen. Treatment regimens will also be compared graphically. All safety data available through the data cutoff will be analyzed and provided.

21.1. Safety Review Committee

Primary oversight for subject safety will be the responsibility of the investigators and sponsor. Additional advisory oversight of subject safety will be provided by an independent Safety Review Committee (SRC) composed of individuals with pertinent medical expertise, who will serve in advisory capacity to the sponsor to provide an additional level of oversight to minimize the chance that clinical study subjects are exposed to unreasonable or unnecessary risks. To enable the assessment of risk-benefit of pegzilarginase during the study, members of the independent SRC will have access to unblinded safety data.

SRC meetings will take place over the course of the study and unblinded TFLs will be produced and provided to the SRC. Details of the SRC unblinded interim analyses are outlined in the safety data reporting plan.

22. CHANGES TO THE ANALYSIS PLAN

22.1. Changes to Protocol-Stated Analyses

By the time the last evaluable subject has completed the placebo-controlled DB portion of the study and the Week 24 data cut is taken, a group of subjects will have data available beyond 24 weeks. Thus, in addition to the 24-week analyses defined in the protocol, select efficacy analyses will be performed for all available data for subjects who have data through 48 weeks (i.e., LTE24).

Regarding the analysis of the primary endpoint, change from Baseline in plasma arginine, the protocol states that 2 separate sensitivity analyses will be performed: 1) comparing the change from Baseline in arginine using a Wilcoxon rank sum test, 2) performing a Wilcoxon rank sum test where those subjects who drop out for treatment-related AEs, withdrawal of consent, or major protocol deviations will receive lower rank than those who complete all 24 weeks or who drop out for reasons clearly not related to study treatment. This second sensitivity analysis was to assess the impact of dropouts. However, as the study has progressed it is observed that there will be very few if any subjects failing to receive all 24 doses. Thus, this second sensitivity analysis is deemed unnecessary.

22.2. Changes to the Statistical Analysis Plan

All versions of the SAP are listed in Section 1. Summaries of changes for each version are described in the subsequent sections.

22.2.1. Version 2.0

The primary purpose for SAP Version 3.0 was to update some analyses depending on data availability and the need to provide particular summaries, as well as to update the analysis periods defined in the outputs.

Updates are summarized below.

1. **Whole Document**

   Description of change: Some planned tables were removed. Listings will still be generated.

   Rationale for change: Following a blinded Dry Run of the data it was determined that some of the planned outputs do not have enough subjects for the purpose of summarizing in a table, or that the tables were deemed unnecessary and that listings will suffice.

2. **Whole document**

   Description of change: The definitions of analysis periods were updated.

   Rationale for change: To better align with the intent of the study and to better clarify the descriptions of the DB and LTE periods, both in the text and in the outputs.

3. **Composite Clinical Outcome: 2MWT, GMFM Part D, and GMFM Part E**

   Description of change: The frequencies of subjects achieving response was updated to reflect ≥1, ≥2, or 3 of the categories (from 1, 2, 3). Additionally, and additional definition

of responder was included (defined as having an improvement in at least 1 component assessment regardless of worsening in any other component).

Rationale for change: to remove ambiguity in the definition and to provide additional information.

information.

4. **Section 19.1**

   Description of change: A summary of Adverse event during the LTE period were added, as well as a summary of exposure-adjusted incidence rates (EAIRs).

   Rationale for change: To better evaluate and compare AEs and AE rates between the DB and LTE periods.

5. **Section 19.5.2**

   Description of change: The section defining Liver Function Test abnormalities was updated and moved.

   Rationale for change: For Clarity as it was not considered an AESI.

22.2.2. Version 2.0

The primary purpose of SAP Version 2.0 was to better define analyses of the study endpoints and to update the methods based upon interactions with regulatory authorities, as well as to better define the analysis periods and timepoints.

Updates are summarized below.

1. **Section 7, Sample Size and Power Considerations**

   Description of change: Power calculation was updated to reflect the updated primary endpoint of change from baseline in logged plasma arginine.

   Rationale for change: Following correspondence with FDA it was agreed to define primary endpoint and key secondary endpoints using change from baseline.

2. **Section 9, Analysis Periods and Populations; Section 10.8, Timepoints; Section 21, Interim Analyses**

   Description of change: Analysis periods and timepoints were added, and corresponding text was revised as appropriate.

   Rationale for change: To clarify.

3. **Section 11, Disposition of Subjects; Section 12, Demographic and Pretreatment Variables; Section 13.2, Prior and Concomitant Medications; Section 14, Dosing Compliance and Treatment Exposure**

   Description of change: Subject disposition analyses were clarified. History of hyperammonemic episodes was added to demographic variables, and baseline disease characteristics were added and defined. Prior and concomitant medications were clarified. Dosing compliance and treatment exposure were clarified.

Rationale for change: To clarify.

4. **Section 15, Protocol Deviations**

   Description of change: A new section regarding protocol deviations was added.

   Rationale for change: To define protocol deviations and corresponding outputs.

5. **Section 16, Derivation of Efficacy Measures; Section 17, Efficacy Analyses; Section 18, Tertiary Analyses; Section 19, Safety Analyses**

   Description of change: Text was revised to more accurately reflect the endpoints being measured and their corresponding derivations, analyses, and outputs, as applicable. Sub-sections were added, revised, or moved accordingly.

   Rationale for change: To clarify and to address feedback from regulatory authorities.

6. **Section 22, Changes to the Analysis Plan**

   Description of change: Changes to protocol-stated analyses were updated, and changes to the SAP were added.

   Rationale for change: To accurately depict the development of the analysis plan.

7. **Incorporation of administrative changes.** Other minor, administrative changes have been incorporated throughout the SAP and are noted in the redline version.

22.2.3. Version 1.0

SAP Version 1.0 was the original SAP for this study.

23. REFERENCES

Berg MD, Schexnayder SM, Chameides L, et al. Part 13: pediatric basic life support: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2010; 122(18 Suppl 3): S862-S875.

Centers for Disease Control and Prevention (United States). CDC Growth Charts: United States (2002). Accessed from https://www.cdc.gov/growthcharts/cdc_charts.htm.

Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988; 74(4): 800-802.

Ouyang J, Carroll KJ, Koch G, Li J. Coping with missing data in phase III pivotal registration trials: Tolvaptan in subjects with kidney disease, a case study. Pharm Stat. 2017; 16(4): 250-266.

Ware JE, Snow KK, Kosinski M, Gandek B. SF-36® Health Survey Manual and Interpretation Guide. Boston, MA: New England Medical Center, The Health Institute; 1993.