Statistical Analysis Plan

NCT Number: NCT03771898

Title: A Global, Multicenter, Single-arm, Matched External Control Study of Intrathecal SHP611 in Subjects with Late Infantile Metachromatic Leukodystrophy

Study Number: SHP611-201

Document Version and Date: Version 2.0, 14 March 2023

Certain information within this document has been redacted (ie, specific content is masked irreversibly from view) to protect either personally identifiable information or company confidential information.

STATISTICAL ANALYSIS PLAN

Study Number: SHP611-201

Study Title: A Global, Multicenter, Single-arm, Matched External Control Study of Intrathecal SHP611 in Subjects with Late Infantile Metachromatic Leukodystrophy

Phase: 2

Version: 2.0

Date: 14-Mar-2023

Prepared by: [redacted]

Based on:

Protocol Version: Amendment 5

Protocol Date: 12-Oct-2022

REVISION HISTORY

Version	Approval Date	Primary Rationale for Revision
0.1 (draft version sent to the FDA as Appendix D of the Type C Briefing Package for the Type C meeting on 20 July 2022)	25-May-2022	[Not Applicable]
1.0	13-Dec-2022	• Incorporation of SHP611-201 Protocol Amendment 5 changes • Addition of FDA recommended analyses • Incorporation of dry-run feedback
2.0	14-Mar-2023	• Pre-specification of the use of centrally rated GMFC-MLD category for all primary analyses, and the use of local or originally adjudicated GMFC-MLD category at Screening for sensitivity analysis • Addition of Baseline definition for GMFM-88 • Addition of details for the quantification of uncertainty in survival probability estimation • Addition of details for the sensitivity analyses on propensity score matching and weighting • Incorporation of dry-run feedback on details of data presentation

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

ABBREVIATIONS

AE	adverse event
AESI	adverse event of special interest
ANCOVA	analysis of covariance
ANOVA	analysis of variance
BLQ	below the lower limit of quantitation
CMH	Cochran-Mantel-Haenszel
COVID-19	coronavirus disease 2019
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee
IRC	Internal Review Committee
ECG	electrocardiogram
eCRF	electronic case report form
FAS	full analysis set
ITT	intention-to-treat
KM	Kaplan-Meier
LLN	lower limit of normal
LOCF	last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
mFAS	modified full analysis set
PD	pharmacodynamic
PK	pharmacokinetic
PPAS	per-protocol analysis set
PRO	patient-reported outcomes
PT	Preferred Term (MedDRA)
SAE	serious adverse event
SAP	statistical analysis plan
SOC	System Organ Class
TEAE	treatment-emergent adverse event

1. OBJECTIVES, ENDPOINTS AND ESTIMANDS

1.1 Study objectives and endpoints

The study objectives and endpoints are summarized in Table 1.

Table 1 Objectives and Endpoints

Objective	Endpoint(s)
Primary	
1. The primary objective of this study is to evaluate the effects of intrathecal (IT) administration of SHP611 (also known as TAK-611) on the time to loss of locomotion, as indicated by category 5 or higher in the Gross Motor Function Classification in Metachromatic Leukodystrophy (GMFC-MLD) compared with matched external control group data in children with late infantile MLD	1. The primary efficacy endpoint is time to loss of locomotion, measured by progression to GMFC-MLD category 5 or higher, or death, whichever occurs first, up to Week 106, evaluated on subjects in Group A
Secondary	
1. To evaluate the effects of IT administration of SHP611 on subjects who experience decline in gross motor function as indicated by GMFC-MLD category 5 or higher, compared with matched external control group data in children with MLD	1. Response in Group A, defined as maintenance of gross motor function at Week 106, evaluated as subjects who do not experience any event within Week 106, where event is defined as a decline in GMFC-MLD to category 5 or higher, or death
2. To evaluate the effects of IT administration of SHP611 on the decline in gross motor function, as measured by an unreversed decline in GMFC-MLD of more than 2 categories compared with matched external control group data in children with MLD, time course of declining gross motor function using the GMFC-MLD, and change from baseline of gross motor function, using the GMFC-MLD	2. Decline in gross motor function using GMFC-MLD: a) Change from baseline at Week 106 and EOS in gross motor function, using the GMFC-MLD b) Decline in gross motor function using GMFC-MLD: Subjects with unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2-categories that has not reverted to a 2-category decline (or better) at Week 106, evaluated on subjects in Group A c) Decline in gross motor function using GMFC-MLD: Time to unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2 categories that has not reverted to a 2-category decline (or better) as of the last recorded observation
3. To evaluate the effects of IT administration of SHP611 on CSF sulfatides (PD biomarker)	3. Change from baseline at Week 106 and EOS in CSF sulfatides levels
4. To evaluate the effects of IT administration of SHP611 on gross motor function, using the Gross Motor Function Measure 88 (GMFM-88) total score in children with MLD	4. Response in Group A, defined as maintenance of gross motor function at Week 106, defined as a GMFM-88 total score ≥40

Table 1 Objectives and Endpoints

Objective	Endpoint(s)
5) To evaluate the effects of IT administration of SHP611 on the time course of declining gross motor function using GMFM-88, as measured by a) an unreversed decline from baseline in GMFM-88 total score of >20 points or unreversed decline to <40 points, whichever occurs first, b) change from baseline of gross motor function, using the GMFM-88 total score, and c) GMFM-88 total score decline of no more than 20 points from baseline and a total score that is ≥40	5) Decline in gross motor function using GMFM-88: a) Time to unreversed decline from baseline at Week 106 and EOS in GMFM-88 total score decrease of >20 points or unreversed decline to a score <40 points, whichever occurs first b) Decline in gross motor function using GMFM-88: Change from baseline at Week 106 and EOS in gross motor function, using the GMFM-88 total score c) Decline in gross motor function using GMFM-88: Subjects in Group A with GMFM-88 total score decrease of ≤20 points from baseline and a total score that is ≥40 at Week 106 and EOS
6) To evaluate the effects of IT administration of SHP611 on expressive language using the Expressive Language Function Classification (ELFC-MLD)	6) Change from baseline at Week 106 and EOS in expressive language using the ELFC-MLD
Pharmacokinetics	
1) To evaluate the concentrations of SHP611 in CSF following single and repeat IT dosing of SHP611	1) CSF parameters: a) Predose concentrations (C_{trough}) of SHP611 at Weeks 0, 5, 9, 13, 26, 40, 53, 79, and 106 b) Postdose concentrations of SHP611 at 6 and 24 hours (Weeks 0 and 106)
2) To evaluate the concentrations and PK parameters of SHP611 in serum following single and repeat IT dosing of SHP611	2) Serum parameters: a) PK parameters after the first dose (Week 0) and after repeated doses (Week 106) of SHP611 determined by noncompartmental analysis will include but not limited to area under the curve (AUC), maximum concentration (C_{max}), and clearance after IT administration (CL/F) b) Predose concentrations (C_{trough}) of SHP611 at Weeks 0, 13, 26, 40, 53, 79, and 106
Safety	
1) Occurrence of treatment-emergent adverse events (TEAEs)	1) Treatment-emergent adverse events (TEAEs)
2) Clinical laboratory testing (serum chemistry, hematology, and urinalysis) and vital signs	2) Changes from baseline at Week 106 and EOS in clinical laboratory testing (serum chemistry, hematology, and urinalysis)
3) Physical examination including documentation of signs and symptoms of MLD and Developmental Questionnaire	3) Change from baseline at Week 106 and EOS in physical examination including documentation of signs and symptoms of MLD (tone, reflexes, and vision)

Table 1 Objectives and Endpoints

Objective	Endpoint(s)
4) 12-lead electrocardiogram (ECG)	4) Change from baseline at Week 106 and EOS in 12-lead electrocardiogram (ECG)
5) CSF laboratory parameters (chemistries and cell counts)	5) Change from baseline at Week 106 and EOS in CSF laboratory parameters (chemistries, cell counts)
6) Development of anti-SHP611 antibodies in CSF and serum	6) Anti-SHP611 antibody responses in CSF and serum at Week 106 and EOS
7) SOPH-A-PORT® Mini S device in subjects with MLD	7) SOPH-A-PORT Mini S assessments will be evaluated using assessments of device implantation, device function, device longevity, and adverse events (AEs) associated with the implant surgery or device
Exploratory	
Objective	**Endpoint(s)**
To evaluate the effects of administration of IT SHP611 on:	**Change in:**
1) CSF, serum and urine biomarkers	1) CSF, serum, and urine biomarkers over time
2) Proton magnetic resonance spectroscopy (MRS) of the brain, specifically N-acetylaspartate/Creatine (NAA/Cr) in white matter	2) MRS metabolite levels specifically: N-acetylaspartate/Creatine over time
3) Eichler MLD MRI severity score	3) Eichler MLD MRI severity score over time
4) Severity score as measured by magnetic resonance imaging (MRI) of the brain	4) Total MLD severity score based on brain MRI over time
5) Volumetric analysis based on MRI of the brain	5) Volumetric analysis of the brain based on MRI over time
6) Global impression of motor function – change (GIMF-C)	6) GIMF-C over time
7) Global impression of motor function – severity (GIMF-S)	7) GIMF-S over time
8) Caregiver burden and subject's health-related quality of life impact in children with MLD by evaluating: a) Caregiver burden as assessed by the Caregiver Impact Questionnaire (CIQ) b) Health Related Quality of Life (HRQOL) as assessed by the Infant Toddler Quality of Life Questionnaire – 97 items (ITQOL-97)	8) Caregiver burden and subject's health-related quality of life: a) Descriptive statistics of the Caregiver Impact Questionnaire (CIQ) item responses over time to inform scoring b) Caregiver burden and subject's health-related quality of life: change in each of the parent and infant/toddler concepts as assessed by the Infant Toddler Quality of Life Questionnaire – 97 items (ITQOL-97) over time

Table 1 Objectives and Endpoints

Objective	Endpoint(s)
9) Healthcare Utilization as measured by the Health Care Utilization Questionnaire (HCUQ)	9) Incidence of hospitalizations, number of days in hospital, reason for admission, and frequency of selected MLD-related procedures (use of feeding tube, use of intubation, and type of respiratory support) over time; total number of additional hospitalizations during the 2-year follow-up
10) Caregiver work productivity and activity impairment as assessed using the Work Productivity and Activity Impairment Questionnaire (WPAI): Specific Health Problem V2.0	10) Caregiver work productivity and activity impairment as assessed using the Work Productivity and Activity Impairment Questionnaire (WPAI): Specific Health Problem V2.0 over time
11) Ability to eat and drink as assessed using the Eating and Drinking Ability Classification System (EDACS) assessments	11) Ability to eat and drink as assessed using the Eating and Drinking Ability Classification System (EDACS) assessments over time

1.2 Estimand(s)

The primary and key secondary estimands are described in Table 2.

Table 2 Estimand Framework

Estimand: [Primary]

Attributes				
Treatment	**Population**	**Variable (or Endpoint)**	**Strategy for Addressing Intercurrent Event (ICE)**	**Population-Level Summary**
150 mg weekly IT administration of SHP611 is the sole intervention.	SHP611-201 enrolled subjects in Group A and matched external control of untreated MLD subjects from the ongoing Global Leukodystrophy Initiative (GLIA-MLD) natural history study.	Time to loss of locomotion, measured by progression to GMFC-MLD category 5 or higher, or death, whichever occurs first, up to Week 106.	A composite strategy for the intercurrent events is considered: **Death after the time-to-event (TTE) starting point** - In this situation, the subject will be regarded as having experienced an event with the corresponding TTE incorporated in the estimation of treatment effect. **Treatment discontinuation** - If the subject discontinued early from the study during the primary treatment period without an observed event, this subject will be right censored at the last GMFC-MLD assessment time point (i.e., if GMFC-MLD category is no more than 4 and the subject is known to be alive).	Difference at Week 106 in survival functions of time to loss of locomotion (measured by progression to GMFC-MLD category 5 or higher, or death, whichever occurs first up to Week 106), between Group A subjects in SHP611-201 and matched external control subjects, as quantified using a weighted average of proportion of patients not reaching the event of interest, with weights derived from the relative size of treated and control units in the strata used for the stratified log-rank test in the primary analysis[a] .

Table 2 Estimand Framework

Estimand: [Secondary # 1 (Efficacy)]				
Attributes				
Treatment	**Population**	**Variable (or Endpoint)**	**Strategy for Addressing Intercurrent Event (ICE)**	**Population-Level Summary**
150 mg weekly IT administration of SHP611 is the sole intervention.	SHP611-201 enrolled subjects in Group A and matched external control of untreated MLD subjects from the ongoing Global Leukodystrophy Initiative (GLIA-MLD) natural history study.	Maintenance of gross motor function at Week 106, evaluated as subjects who do not experience any event within Week 106, where event is defined as a decline in GMFC-MLD to category 5 or higher, or death.	A composite strategy is considered using the worst outcome for ICE of death after the TTE starting point, or ICE of treatment discontinuation. The following intercurrent events are considered: **Death after the TTE starting point -** In this situation, "no maintenance of gross motor function" will be used as outcome of the endpoint **Treatment discontinuation -** If the subject discontinued early from the study during the primary treatment period in SHP611-201, "no maintenance of gross motor function" will be used as outcome of the endpoint.	Difference at Week 106 in proportions of subjects who maintained gross motor function, evaluated as subjects who do not experience any event within Week 106 (where event is defined as a decline in GMFC-MLD to category 5 or higher, or death), between Group A subjects in SHP611-201 and matched external control subjects, and the corresponding two-sided CI[b].
Estimand: [Secondary # 2 (Efficacy)]				
Attributes				
Treatment	**Population**	**Variable (or Endpoint)**	**Strategy for Addressing Intercurrent Event (ICE)**	**Population-Level Summary**
150 mg weekly IT administration of SHP611 is the sole intervention.	SHP611-201 enrolled subjects in Group A and matched external control of untreated MLD subjects from the ongoing Global Leukodystrophy Initiative (GLIA-MLD) natural history study.	Unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2-categories that has not reverted to a 2-category decline (or better) at Week 106.	A composite strategy is considered using the worst outcome for ICE of death after the TTE starting point, or ICE of treatment discontinuation. The following intercurrent events are considered: **Death after the TTE starting point** - In this situation, "subject achieved unreversed decline from baseline in GMFC-MLD of more than 2 categories" will be used as outcome of the endpoint. **Treatment discontinuation** - If the subject discontinued early from the study during the primary treatment period in SHP611-201, "subject achieved unreversed decline from baseline in GMFC-MLD of more than 2 categories" will be used as outcome of the endpoint.	Difference in proportions of subjects who do not experience unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2-categories that has not reverted to a 2-category decline (or better) at Week 106, between Group A subjects in SHP611-201 and matched external control subjects, and the corresponding two-sided CI[b].

Table 2 Estimand Framework

[a] Interval censoring methods will be used, with event assumed to have first happened between the last visit/encounter prior to the event observation, and the visit/encounter when the event is first observed. Kaplan-Meier survival curves suitable for interval censored data will be presented. Non-parametric estimates of the probability of not experiencing the event in the appropriate intervals (for both GLIA-MLD and SHP611-201) will be presented. See Section 7.5.1.2 for more details.
[b] See Section 7.1.3 and Section 7.5.2 for the calculation of the two-sided confidence interval and assessment of efficacy comparison.

2. STUDY DESIGN

2.1 SHP611-201 Study Design

SHP611-201 is a single-arm, matched external control, global, multicenter, Phase 2 trial. The study was planned to enroll up to 42 subjects with late infantile MLD who have an initial onset of neurological symptoms documented prior to 30 months of age (Groups A, B, C, and F), or who are minimally symptomatic and ≥6 to <18 months of age (Group D), or who are early symptomatic and ≥12 to <18 months of age (Group E). Minimally symptomatic is defined as being without clear symptoms of MLD or only showing mild symptoms (such as weakness) that do not meet the criteria for a GMFC-MLD category of >0 (NB: no subjects were enrolled in Groups D or E). The rate and severity of disease progression is well documented in late infantile MLD (Kehrer et al. 2011b; Kehrer et al. 2011a). A distinguishing feature of the definition of late infantile MLD is the early age at disease symptom onset with a majority of patients with late infantile MLD showing first motor dysfunction before the age of 18 months. Six subject groups are defined for this study based on age and motor dysfunction at screening:

- Group A (GMFC-MLD category 1 or 2): Approximately 16 subjects who are 18 to 48 months of age with a GMFC-MLD category of 1 or 2
- Group B (GMFC-MLD category 3): Up to 8 subjects who are 18 to 72 months of age with a GMFC-MLD category of 3
- Group C (GMFC-MLD category 4): Up to 8 subjects who are 18 to 72 months of age with a GMFC-MLD category of 4
- Group D (minimally symptomatic): Up to 3 subjects who are ≥6 to <18 months of age, with the same ASA allelic constitution as an older sibling with confirmed late infantile or juvenile onset MLD
- Group E (GMFC-MLD category 1 or 2, <18 months of age): Up to 3 subjects who are ≥12 to <18 months of age, with documented diagnosis of MLD per inclusion criteria 1 and 2 who have achieved stable walking (defined as at least 1 month of independent walking) and who have a GMFC-MLD category of 1 or 2
- Group F (GMFC-MLD category 5 or 6): Up to 4 subjects who are 18 to 72 months of age with a GMFC-MLD category of 5 or 6

GMFC-MLD is a validated categorical scoring system with 7 levels, with 0 defining no abnormalities in gross motor function, to 6, defining loss of all gross motor function, including head control:

Table 3 Gross Motor Function Classification in MLD (GMFC-MLD) Scale

GMFC-MLD Category	Description
0 (M0)*	Walking without support with quality of performance normal for age
1 (M1)	Walking without support but with reduced quality of performance, that is, instability when standing or walking
2 (M2)	Walking with support. Walking without support not possible (fewer than five steps)
3 (M3)	Walking with or without support not possible. Sitting without support and locomotion such as crawling or rolling still present.
4 (M4)	(a) Sitting without support but no locomotion, or (b) Sitting without support not possible, but locomotion such as crawling or rolling
5 (M5)*	Neither locomotion nor unsupported sitting are possible, but head control is present
6 (M6)	Loss of any locomotion as well as loss of any head and trunk control

* In the GLIA-MLD study, the GMFC-MLD scale includes two additional subcategories: M0b (ambulation present, but of unknown quality) and M5b (loss of locomotion, inability to sit unassisted, but quality of head control unknown), created to score the GMFC-MLD from retrospective medical chart review when there is insufficient detail in the medical provider notes for the rater to differentiate M0-M2 (GMFC-MLD category 0-2) or M5-M6 (GMFC-MLD category 5-6).

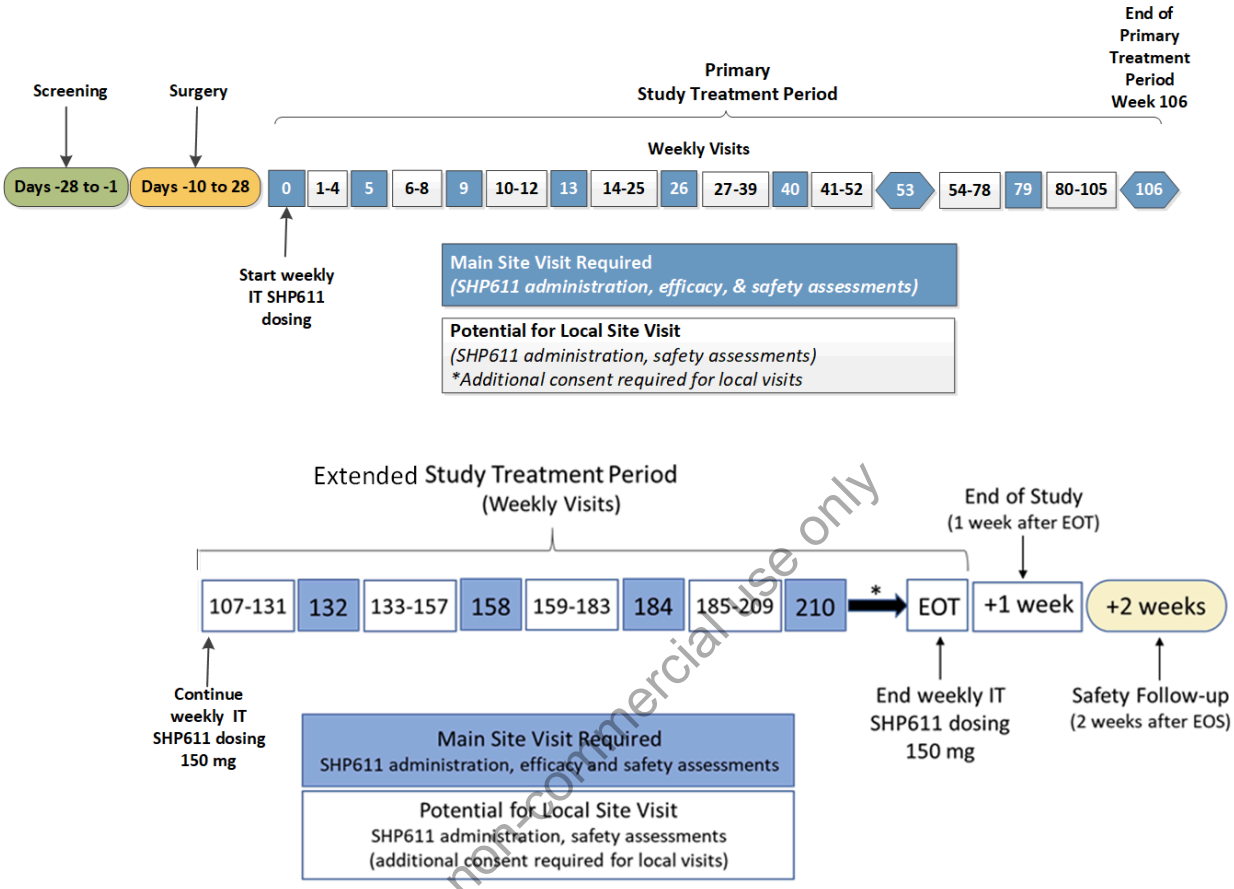
Source: Kehrer, C., Blumenstock, G., Raabe, C. & Krageloh-Mann, I. 2011a. Development and reliability of a classification system for gross motor function in children with metachromatic leucodystrophy. *Dev Med Child Neurol,* 53, 156-60.

Subjects weighing ≥7 kg (15.4 lbs) will receive 150 mg IT SHP611 weekly. It is anticipated that the majority of subjects will receive 150 mg IT weekly for a total treatment duration of 105 weeks; however, subjects weighing ≥5 kg (11.0 lbs) to <7 kg (15.4 lbs) will receive 100 mg IT SHP611 weekly until they weigh ≥7 kg, at which time they will begin dosing with 150 mg IT SHP611 weekly.

The study will consist of a screening period of up to 28 days. Implantation of the SOPH-A-PORT Mini S IDDD may occur during a period of up to 10 days prior to the first administration of IT SHP611 to 28 days after the first administration of IT SHP611. IT SHP611 administrations that occur prior to implantation of the IDDD will be administered via lumbar puncture (LP).

Schema

Figure 1 Study Schematic Diagram: Primary and Extension Treatment Periods



EOS=end of study; EOT=end of treatment; IT=intrathecal.

*Subjects may continue treatment in the study beyond Week 210. They will continue to follow the same schedule of assessments, ie, weekly dosing and main site visits every 6 months. Subjects will continue treatment until they or their parents/guardians decide to discontinue treatment; the sponsor discontinues the study; the subject is discontinued from the study due to medical or safety concerns; or the product becomes commercially available in the subject's country of residence, whichever comes first.

Subjects will be assessed according to the following schedule:

- Screening (-28 to -1 days)
- Surgical implantation of IDDD (-10 to 28 days)
- Primary treatment period (Week 0 [baseline assessments prior to dosing] through Week 105)
- End of primary treatment period (Week 106)
- Extension treatment period (from Week 106 administration of SHP611)
- End of treatment (EOT) (last administration of SHP611)
- End of study (EOS) (1 week after EOT)
- Safety follow-up (2 weeks after EOS)

After the primary treatment period is completed at Week 106, subjects may participate in the extension period of the study where they may continue to receive treatment with SHP611 for an extended duration of time.

Subjects will receive weekly treatment until they or their parents/guardians decide to discontinue treatment; the sponsor discontinues the study; the subject is discontinued from the study due to medical or safety concerns; or the product becomes commercially available in the subject's country of residence, whichever comes first. During the extension period, main site assessments will be scheduled every 6 months. These assessments may be skipped for a visit at the discretion of the investigator and upon discussion of the investigator with the Medical Monitor if it is determined that the subject is unable to perform the assessments. At the EOT visit, subjects will receive their last administration of SHP611 and comprehensive assessments will be completed at the EOS visit.

The primary treatment period is planned for 106 weeks with an extension period starting at Week 106 administration of SHP611. The planned overall duration of each subject's involvement in the study is approximately 26 months from screening to the last scheduled visit for the primary treatment period with the extension period planned to continue until Mar 2025, or product commercialization date, or until the program is discontinued.

The primary analysis will be performed after all subjects complete the primary treatment period of the study; the database will be locked with the results presented in a clinical study report (NB: **the sponsor will remain blinded to post-baseline efficacy-related data until database lock**). Following completion of the extension period, the database will be locked again and the results of the entire study (primary treatment and extension periods) will be described in the final clinical study report.

3. MATCHED EXTERNAL CONTROL

The efficacy of SHP611 will be evaluated by comparison of SHP611-201 enrolled subjects in Group A with matched external control subjects, i.e., untreated MLD subjects who have received no investigational product or therapy. The external control data will come from the ongoing Global Leukodystrophy Initiative of Metachromatic Leukodystrophy (GLIA-MLD), a retrospective longitudinal natural history study of MLD, under the regulatory umbrella of the Myelin Disorders Biorepository Project (MDBP) and Global Leukodystrophy Initiative Clinical Trial Network (GLIA-CTN) Protocol 14-011236 (December 2016).

For this matched external control study, the optimal full matching method using only the baseline data from SHP611-201 and GLIA-MLD (refer to Section 3.2 for the selection of baseline encounter in GLIA-MLD) will be applied to balance the baseline observed characteristics between the treated and the external control groups. For both the GLIA-MLD and SHP611-201 datasets, any post baseline efficacy data will be blinded to the study team until database lock.

Due to the retrospective scoring nature for GLIA-MLD, multiple encounters (defined as any documented physical or virtual contact between a subject and healthcare practitioner, during which an assessment or clinical activity is performed) could be generated from medical notes by CHOP, and only encounters with available GMFC-MLD assessments will be used in the analyses.

In the GLIA-MLD study, the GMFC-MLD scale included two additional subcategories: M0b (ambulation present, but of unknown quality) and M5b (loss of locomotion, inability to sit unassisted, but quality of head control unknown). These two additional subcategories were introduced to score the GMFC-MLD from retrospective medical chart review when there was insufficient detail in the medical provider notes for the rater to differentiate M0-M2 (GMFC-MLD category 0-2) or M5-M6 (GMFC-MLD category 5-6). These were not necessary for reviewing a video recording of a live examination, as the raters could ascertain the quality of locomotion and head control themselves.

3.1 Matched External Control Subjects for Group A from GLIA-MLD

The matched external control group must have data for at least baseline gross motor function evaluation. Selection of the external control subjects from GLIA-MLD will follow a set of criteria as similar as possible to the inclusion criteria for Group A in the SHP611-201 study protocol.

A filtering process will be applied to select the external control subjects from the GLIA-MLD database, by meeting all of the following 3 filtering criteria:

1. **Filtering criterion 1:** requiring documented diagnosis of MLD, based on
    - low ASA activity in leukocytes AND elevated sulfatides in urine.
      OR
    - biallelic variants in ARSA AND (either low ASA activity in leukocytes OR elevated sulfatides in urine).

2. **Filtering criterion 2**: requiring documented gait disorder. Patients will be considered qualifying if they present with a gait disorder before 2.5 years (30 months) of age and have a medical record reporting a gait abnormality including, but not limited to, the following terms: ataxia, spasticity, and hyper/hypotonia. See Appendix 10.3 for the complete list of terms.
3. **Filtering criterion 3**: subjects will be considered qualifying if they have at least 1 clinical encounter occurring between the age of 18 to 48 months with a GMFC-MLD category either 1 or 2.

For the filtering criterion 2, age at presentation of gait disorder is required to be less than 2.5 years (30 months), which is defined as the minimum age obtained from the following three variables:

- age when MLD was first noted as a suspected diagnosis
- age at presentation by caregivers
- age at presentation by physician

All of the information above is collected through medical documentation of parent/guardian report or by physicians, available in the medical charts.

In addition, for the filtering criterion 2, in terms of the symptoms noted, patients will be required to have met at least one of the following criteria:

- Have a GMFC-MLD category M1 or above from one encounter which occurred before 30 months
- Have a documented motor finding before 30 months, such as spasticity, truncal hypotonia or gait/truncal ataxia including the complete list of terms in Appendix 10.3

For the filtering criterion 3, subjects will be considered qualifying if they have at least 1 clinical encounter occurring between the age of 18 to 48 months with a GMFC-MLD category either M1 or M2. Note that for this criterion, encounters with GMFC-MLD category of M0b (ambulation present, but of unknown quality) are NOT considered, unless it is the only possible 'qualifying encounter'.

This additional subcategory is given to indicate insufficient detail in the medical provider's notes for the rater to differentiate M0-M2 (GMFC-MLD category 0-2) during retrospective medical chart review.

In addition to requiring the ability to ascertain GMFC-MLD category and age for the first clinical encounter (i.e., a GMFC MLD category either M1 or M2 when the subject was between 18-48 months of age), this encounter should occur before any therapeutic intervention (TI), where TI is defined as bone marrow transplant, gene therapy or enzyme replacement therapy.

3.2 Baseline Encounter in GLIA-MLD

To select the TTE starting point (index date) in the external control group from GLIA-MLD, if a subject has multiple 'qualifying encounters' (i.e. an encounter that meets the filtering criteria for 'external control subjects for Group A'), then the encounter that minimizes the mean adjusted multidimensional Euclidean distance (based on age at MLD symptom onset, and duration from onset to the qualified encounter) to Study SHP611-201 Group A subjects at the treatment initiation will be used as the index date, with details provided below.

By applying filtering criterion 3 as above, a GLIA-MLD subject needs to have at least one clinical encounter between the age of 18 to 48 months with the GMFC-MLD category of either M1 or M2, though some subjects may have more than one such 'qualifying encounter'. All such encounters could be eligible candidates for the TTE 'starting point'. To determine the TTE starting point in GLIA-MLD, an objective selection process will be applied for each subject in GLIA-MLD who has qualified as a 'external control for Group A' subject with multiple 'qualifying encounters':

1. Calculate the mean adjusted multidimensional Euclidean distance between each qualifying encounter for a GLIA-MLD external control subject and all Group A subjects with the same GMFC-MLD category at the Study 201 Screening visit in Study SHP611-201. Here the distance is based on the 2 continuous covariates including:
   i) age at MLD symptom onset, and
   ii) duration from onset of MLD symptoms to the qualified encounter.

   Both covariates will be adjusted for the respective standard deviation so that the distance is not influenced by variance or unit of measurement.
2. The encounter associated with the lowest distance is selected as the TTE starting point for the GLIA-MLD subject.

Details of the mathematical formulation is provided in Appendix 10.6.

For subjects in GLIA-MLD with the GMFC-MLD category of M0b at an encounter that would otherwise qualify as potential TTE starting point, the encounter with the M0b category will be used only if the patient does not have any 'qualifying encounters' with a GMFC-MLD score of M1 or M2 (i.e., abnormal ambulation) on subsequent examination clearly recorded. In other words, the M0b encounter will be used if it is the only possible 'qualifying encounter', and it will be treated as M1. For GLIA-MLD subjects with any documented GMFC-MLD score of category M1 or M2 within the proper age range of 18 to 48 months, any previous encounters in GLIA-MLD scored as M0b will be excluded. This approach will ensure that all potentially eligible controls are included, while encounters with more reliable assessments are used in the matching process.

3.3 Optimal Full Matching

The optimal full matching is set to be conducted prior to unblinding of any post-baseline efficacy data. Group A subjects from SHP611-201 and external control subjects from GLIA-MLD will be matched based on GMFC-MLD score (GMFC-MLD category at the Screening visit and at the TTE starting point will be used for SHP611-201 and GLIA-MLD subjects, respectively), age at MLD symptom onset, and duration from MLD symptom onset to TTE starting point using optimal full matching process described in the following steps:

1) Exact matching on GMFC-MLD score will be used, which requires the SHP611-201 subject and its matched external control subject from GLIA-MLD to have the same value of the GMFC-MLD category (i.e., M1 and M2). The GMFC-MLD category at the Screening visit and at the TTE starting point (Section 3.2) will be used for SHP611-201 and GLIA-MLD subjects, respectively.
2) The matching process is comprised of optimal full matching (Austin and Stuart 2015), minimizing Mahalanobis distance (based on the 2 covariates of age at MLD symptom onset and duration from MLD symptom onset to TTE starting point) with exact matching on GMFC-MLD category (M1 or M2) as described in Step 1 above. The average treatment effect for the treated (ATT) weight will be used. Optimal full matching divides all treated and external control subjects in the analysis population into a collection of 1:n and m:1 strata. See SAS sample code in Appendix 10.4 for more details.
3) Matching diagnostics will be conducted. External control subjects that affect the quality of matching severely will be excluded from the control group before unblinding of post-baseline data (i.e., 'pruning') and documented.

The subjects identified through the above process will be considered as the "matched external control subjects for Group A" from GLIA-MLD (also referred to as the GLIA-MLD matched external control group) and will be used in the primary analysis.

3.3.1 Matching Diagnostics

For the matching diagnostics, the matched sample as well as the original unmatched sample will be listed by treatment groups (via listings), accompanied with all corresponding baseline variables used in matching process.

A listing of all baseline covariates will be presented with summary statistics separately for SHP611-201 and GLIA-MLD, for all observations (i.e., the original unmatched sample) and for observations in the matched samples by strata.

To evaluate the degree of cohort balance achieved through matching, the standardized mean difference between the treated and external control group for each covariate will be calculated. In addition, the ratio of the variances of the SHP611-201 group to the matched control group post-balancing will be reported for each variable. In the scenario of unacceptable cohort balance between the SHP611-201 and external control group across covariates, other cohort balancing approaches may be explored prior to unblinding of post-baseline data and will be documented.

In addition, weighted plots will be presented to assess variable differences between the SHP611-201 group and the matched external control group for all variables, with weights derived from the respective matching or weighting process. These plots include bar charts for categorical variables, box plots for continuous variables, and standardized differences plots for continuous variables. Q-Q plots of continuous variables from two groups before and after matching will also be used to assess group balance.

Any external control subjects that affect the quality of matching severely, as identified by the above matching diagnostics (e.g. the standardized differences plots or Q-Q plots), will be excluded from the external control group. After the removal, the remaining sample will go through the matching process again. The entire matching process including any removal of any control subjects will be documented prior to unblinding of post-baseline data.

A sensitivity analysis is also planned in Section 7.5.1.3 that does not exclude the aforementioned external control subjects that affect the quality of matching severely.

4. STATISTICAL HYPOTHESES AND DECISION RULES

4.1 Statistical Hypotheses

The primary efficacy of SHP611 will be evaluated by comparison of SHP611-201 enrolled subjects in Group A with matched external control subjects in GLIA-MLD. For the primary efficacy endpoint, the hypothesis that SHP611 treatment is effective in late-infantile MLD will be tested by assessing whether it delays decline in gross motor function in comparison with the external control group (the alternative hypothesis).

4.2 Statistical Decision Rules

Not applicable.

4.3 Multiplicity Adjustment

In order to protect the study-wide type I error at the 1-sided 0.025 level for testing the primary and secondary hypotheses, the Fixed-Sequence Test procedure will be applied. Specifically, the testing will be conducted in the following order using the modified Full Analysis Set (mFAS, as defined in Section 6.2):

1. The primary efficacy endpoint of time to loss of locomotion, measured by progression to GMFC-MLD category 5 or higher, or death, whichever occurs first, up to Week 106, evaluated on subjects in Group A.
2. The secondary efficacy endpoint of response in Group A, defined as maintenance of gross motor function at Week 106, evaluated as subjects who do not experience any event within Week 106, where event is defined as a decline in GMFC-MLD to category 5 or higher, or death.
3. The secondary efficacy endpoint of unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2-categories that has not reverted to a 2-category decline (or better) at Week 106, evaluated on subjects in Group A.

A subsequent test for the secondary endpoints can only be reported as significant if all prior 1-sided tests are also found significant at the 0.025 level of significance. If prior 1-sided tests are not found to be statistically significant, 1-sided p-values generated for latter analyses will be described as nominally significant if less than or equal to 0.025. Multiplicity is not adjusted for other endpoints in this study.

5. SAMPLE-SIZE DETERMINATION

Per original Protocol, sample size was calculated based on the response in Group A, (i.e., the maintenance of gross motor function at Week 106, evaluated as no greater than 2 categories decline from baseline in GMFC-MLD), and at least 12 paired completers is required to detect a treatment difference for a desired power of 90%, using McNemar's test at a 2-sided significance level of 0.05, with the assumption that the response rates in Group A of SHP611-201 study and the GLIA-MLD matched external control group are 65% and 10% respectively. Furthermore, to adjust for potential unmatched and early discontinuation of subjects, a total of 16 subjects are originally planned to be enrolled into Group A of the current study. As the GLIA-MLD external control group for Group A cohort was expected to be large, efficient matching was plausible and incorporated in the assumptions.

Type I error and power were assessed for the time to event primary endpoint for sample sizes similar to that planned in the original protocol, through simulations using an interval censoring approach. Comparable assumptions were made on the response rates for the time to event primary endpoint in simulations. The response rates (i.e., proportion of subjects not reaching GMFC-MLD category 5 or higher, or death) at Week 106 for Group A in the current study and the GLIA-MLD matched external control group are 65% and 10%, respectively. The encounter structure from a subset of the external control subjects for Group A from the GLIA-MLD natural history study was considered. A simulated event time was randomly matched to a GLIA-MLD encounter schedule and was considered censored if it could not be observed within 2 years under the matched schedule. The event time was assumed to follow a Weibull distribution, with a range of compatible shape parameters that allow approximation of the target response rates of 65% and 10% in the two groups, and a 25% censoring proportion for the GLIA-MLD matched external control group. The type I error was preserved well. The power is assessed to be approximately between 71% to 82%. Matching efficiency is not assumed in the simulations.

6. ANALYSIS SETS

6.1 Screened Set

The Screened Set will consist of all subjects from Study SHP611-201 who have signed informed consent.

6.2 Safety Analysis Set

The Safety Analysis Set will consist of all subjects from Study SHP611-201 (Groups A-F) who receive at least 1 dose of SHP611, or subjects who have undergone the IDDD implantation procedure.

6.3 Full Analysis Set

The Full Analysis Set (FAS) will consist of all subjects from the SHP611-201 Safety Analysis Set who receive at least 1 dose of SHP611 and have at least a screening GMFC-MLD assessment.

6.4 Modified Full Analysis Set

The modified Full Analysis Set (mFAS) will consist of all subjects from Group A in the FAS and the matched external control subjects for Group A from GLIA-MLD natural history study obtained after matching as described in Section 3.3.

6.5 Pharmacokinetic Analysis Set

The Pharmacokinetic (PK) Analyses Set will consist of all subjects from the SHP611-201 Safety Analysis Set who receive at least 1 dose of SHP611 and have at least 1 post-dose measurable [i.e., not below quantifiable limits (BQL)] concentration of SHP611 in serum or CSF.

6.6 Immunogenicity Analysis Set

The Immunogenicity Analyses Set will consist of all subjects from the SHP611-201 Safety Analysis Set who receive at least 1 dose of SHP611 and have at least 1 anti-SHP611 antibody assessment with reportable result in serum or CSF.

7. STATISTICAL ANALYSIS

7.1 General Considerations

All inferential efficacy analyses will be based on the mFAS. All statistical tests will be 1-sided hypothesis tests performed at the 0.025 level of significance. All confidence intervals will be 2-sided 95% confidence intervals, unless stated otherwise. Sensitivity analyses of the primary and key secondary endpoints will also be based on the mFAS.

Descriptive analysis of efficacy endpoints for all groups (A-F) will be conducted over the FAS.

Exploratory statistical and graphical evaluations involving Groups B, C, D, E, and F may be conducted.

The centrally rated GMFC-MLD category (adjudicated as needed) for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 5.0, will be used for all corresponding inferential and descriptive analyses. This centrally rated GMFC-MLD category at Screening will also be used to define SHP611-201 subject groups A-F (Section 2.1), select the baseline encounter in GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3).

The definition of Baseline is provided in Appendix 10.2, and it will be used for all change from Baseline analyses.

Where applicable, variables will be summarized descriptively by study visit.

7.1.1 Handling of Treatment Misallocations

Not applicable.

7.1.2 Analysis Approach for Continuous Variables

For continuous variables, descriptive statistics will include the number of subjects with non-missing values, mean, median, standard deviation (SD), minimum, and maximum values, unless specified otherwise. These will be tabulated by subject group and overall. Means and medians will be presented to 1 more decimal place than the recorded data. The SDs will be presented to 2 more decimal places than the recorded data.

Where applicable, analysis of covariance, utilizing the corresponding baseline levels as covariate, will be utilized to assess the treatment effect for the continuous endpoints. Two-sided 95% Confidence Intervals will be constructed for the difference between the two comparator arms.

7.1.3 Analysis Approach for Binary Variables

For the binary variables, descriptive statistics will include the counts and proportions of each value, including a missing category if applicable, unless specified otherwise. These will be tabulated by subject group and overall. The denominator for the proportion will be based on the number of subjects who provided non-missing responses to the binary variable.

Where applicable, confidence intervals of two-sided 95% coverage will be constructed for the proportions for each comparator arm using the Clopper-Pearson method, and for the difference between the two comparator arms using the Wilson score method.

7.1.4 Analysis Approach for Categorical Variables

Categorical variables will be summarized by the number and percentage of subjects in each category, including a missing category if applicable. For categorical variables, descriptive statistics will include counts and proportions of each category, unless specified otherwise. These will be tabulated by subject group and overall. The denominator for the proportion will be based on the number of subjects who provided non-missing responses to the categorical variable. For inferential analyses of the selected secondary endpoints, refer to Section 1.2 on Estimand(s).

7.1.5 Analysis Approach for Time-to-Event Variables

All time-to-event variables will be presented with Kaplan-Meier survival curves.

Where applicable, interval or right censoring method will be used for the time to event variables (see each endpoint for specific analysis method).

7.2 Disposition of Subjects

Summary of subject disposition will be presented for the treatment group (SHP611) and the GLIA-MLD matched external control group from GLIA-MLD, respectively.

For the SHP611-201 group, subject disposition will be summarized for the Screened Set by subject group and overall. Subject disposition includes the number of subject counts in the following categories:

- subjects in the Screened Set
- subjects in the Safety Analysis Set
- subjects in the FAS
- subjects who missed no more than two consecutive SHP611 doses due to COVID-19 (during the primary treatment period, during the extension period, and overall, respectively)
- PK analysis set
- Immunogenicity analysis set
- subjects who withdrew from the study and further classified by reasons of withdrawal (during the primary treatment period, during the extension period, and overall, respectively), and

- subjects who are still in the study (applicable for the primary analysis when all subjects complete the primary treatment period) will be presented

All percentages will be based on the number of subjects in the Safety Analysis Set by subject group and overall, respectively.

For the GLIA-MLD matched external control group, the number of subjects meeting each filtering criterion 1-3 (Section 3.1) will be presented through a funneling process (Appendix 10.7) for those that become eligible as external control subjects for Group A. The number of matched control subjects for Group A from GLIA-MLD will also be reported.

7.3 Protocol Deviations

Protocol deviations (PDs) will be recorded. The CRO/Sponsor will classify major/significant and minor/non-significant protocol deviations per the agreed study Deviations Rules Document. The study team will review the protocol deviations and their classification throughout the study.

For any criteria for protocol deviations that can be completely implemented by a computer program, the detailed algorithm will be agreed upon. Details of such algorithms will be included in the derived dataset specifications and finalized before data unblinding. Non-programmable protocol deviations identified by medical monitoring will be incorporated into the database.

Confirmed major and minor protocol deviations will be documented in the Protocol Deviation tracker for the study. Major/minor protocol deviations will be listed and summarized for the Safety Analysis Set by protocol deviation category and site.

7.4 Demographic and Other Baseline Characteristics

7.4.1 Demographics and Baseline Characteristics

A summary of demographics information and patient characteristics will be presented for the treatment group (SHP611) and GLIA-MLD matched external control group, respectively.

SHP611-201

For SHP611-201, a summary of demographics information will be presented for the Safety Analysis Set and FAS by subject group and overall. The demographic and baseline characteristics consist of

- sex
- age (month)
- genotype
- race
- ethnicity
- head circumference (cm)
- baseline weight (kg)
- baseline height (cm)

- baseline body mass index (BMI)
- age at MLD symptom onset
- age at MLD diagnosis
- age at Screening visit
- age at TTE starting point
- duration from age at MLD symptom onset to TTE starting point
- GMFC-MLD score at Screening visit

GLIA-MLD

For GLIA-MLD, a summary of demographics information will be presented for the matched external control subjects for Group A. The demographic information consists of

- sex
- race
- ethnicity
- baseline head circumference (cm)
- baseline weight (kg)
- baseline height (cm)
- age at MLD symptom onset
- age at MLD diagnosis
- age at 'baseline encounter' (defined in Section 3.2)
- duration from age at MLD symptom onset to 'baseline encounter'
- GMFC-MLD score at 'baseline encounter'

will be summarized using descriptive statistics.

Age at MLD diagnosis, will further be summarized by the type of diagnosis:

- Low ASA activity in leukocytes AND elevated sulfatides in urine
- Biallelic variants in ARSA AND (either low ASA activity in leukocytes OR elevated sulfatides in urine)

The following age-related variables, if available, will be summarized descriptively:

- age when MLD was first noted as a suspected diagnosis
- age at presentation by low Arylsulfatase A (ASA) activity in leukocytes
- age at presentation by elevated sulfatides in urine
- age at presentation by molecular diagnosis
- age at presentation by caregivers
- age at presentation by physician

The symptoms noted at time of initial presentation (initial reported symptoms), if available, will be reported and summarized descriptively.

All demographics information and baseline characteristics collected will be reported in by-subject listings for all subject groups in FAS and the GLIA-MLD matched external control group.

7.4.2 Medical History

For SHP611-201, medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 21.0 or higher. Medical history will be summarized by the number and percentage of subjects in the Safety Analysis Set by system organ class (SOC), and preferred term (PT) for each subject group. Medical history will also be listed by subject.

In GLIA-MLD, all the medical records documented in the encounter history (no matter prior/posterior to the 'baseline encounter'), will be listed for all matched external control subjects for Group A, classified by the following domains:

- Vision, Communication, and Language
- General Gross Motor Function
- Gross Motor Function Measures (GMFM-88)
- Eating and Manual Ability
- Behavioral and Social
- Urinary function
- Cognitive Metrics

7.4.3 Prior and Concomitant Medications, Procedures and Therapies

Prior medication (procedures) is defined as any medication (procedures) with the start date and end date prior to the date of the first dose of investigational product in SHP611-201. Concomitant medication is defined as any medication with a start date prior to the date of the first dose of investigational product in SHP611-201 and end date after the first dose of investigational product in SHP611-201 or with a start date between the dates of the first and last doses of investigational product, inclusive. Concomitant procedure is defined as any procedure with a start date between the dates of the first and last doses of investigational product, inclusive. Any medication (procedure) with a start date after the date of the last dose of investigational product will not be considered a concomitant medication (procedure). Prior and concomitant medications include medications administered within 30 days prior to the Screening visit (Day -28 to -1) and through the final study contact (including protocol-defined follow-up period) are regarded. Prior and concomitant medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) dated March 2018. The prior and concomitant medications will be summarized by anatomical therapeutic chemical (ATC) level 4 and PT for each subject group and overall. Multiple medication usage by a subject in the same category will be counted only once.

All summary of prior and concomitant medications (procedures) will be performed using the Safety Analysis Set by SHP611-201 subject groups (A-F) and overall.

Prior and concomitant medications will be listed for the Safety Analysis Set using verbatim terms and preferred terms (PTs). Prior and concomitant procedures and therapies will also be listed by subject.

Missing or partial medication dates will not be imputed in the database and will stay missing or partial in data listings. However, a conservative imputation approach will be adopted in such cases so that the medication will be deemed to be concomitant if it cannot be definitively categorized to have occurred prior to first dose of study treatment. Similar logic will be applied to deal with missing and partial date for prior and concomitant procedures.

Incomplete start date imputation rules:

The following rules will be applied to impute the missing numerical fields. If the stop date is complete and the imputed start date is after the stop date, then the start date will be imputed using the stop date.

Missing day and month

- If the year of the incomplete start date is the same as the year of the date of the first dose of investigational product, then the day and month of the date of the first dose of investigational product will be assigned to the missing fields
- If the year of the incomplete start date is before the year of the date of the first dose of investigational product, then December 31 will be assigned to the missing fields
- If the year of the incomplete start date is after the year of the date of the first dose of investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete start date are the same as the month and year of the date of the first dose of investigational product, then the day of the date of the first dose of investigational product will be assigned to the missing day
- If either the year is before the year of the date of the first dose of investigational product or if both years are the same but the month is before the month of the date of the first dose of investigational product, then the last day of the month will be assigned to the missing day
- If either the year is after the year of the date of the first dose of investigational product or if both years are the same but the month is after the month of the date of the first dose of investigational product, then the first day of the month will be assigned to the missing day.

Incomplete stop date imputation rules:

The following rules will be applied to impute the missing numerical fields. If the date of the last dose of investigational product is missing, then replace it with the last visit date. If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be equal to the start date.

Missing day and month

- If the year of the incomplete stop date is the same as the year as of the date of the last dose of investigational product, then the day and month of the date of the last dose of investigational product will be assigned to the missing fields
- If the year of the incomplete stop date is before the year of the date of the last dose of investigational product, then 31 December will be assigned to the missing fields
- If the year of the incomplete stop date is after the year of the date of the last dose of investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete stop date are the same as the month and year of the date of the last dose of investigational product, then the day of the date of the last dose of investigational product will be assigned to the missing day
- If either the year is before the year of the date of the last dose of investigational product or if both years are the same but the month is before the month of the date of the last dose of investigational product, then the last day of the month will be assigned to the missing day
- If either the year is after the year of the last dose of investigational product or if both years are the same but the month is after the month of the date of the last dose of investigational product, then the first day of the month will be assigned to the missing day.

7.5 Efficacy Analysis

7.5.1 Primary Endpoint(s) Analysis

7.5.1.1 Derivation of Endpoint

The primary efficacy endpoint is time to loss of locomotion, measured by progression to GMFC-MLD category 5 or higher, or death, whichever occurs first, up to Week 106, evaluated on subjects in Group A.

The TTE starting point in Group A subjects of SHP611-201 is set at the visit where patients receive the first dose of SHP611 (Visit 0). For the control group, the derivation of TTE starting point in GLIA-MLD is described in Section 3.2.

The event of loss of locomotion is defined as a decline in GMFC-MLD from starting point to category 5 or higher, or death, whichever occurs first, up to Week 106 (or 2 years), evaluated on subjects in Group A (or matched external control subjects). If a GMFC-MLD category M5b encounter is identified in a matched control subject, M5b will be considered as part of the event definition for loss of locomotion (M5/M5b/M6 or death), therefore the encounter will be included in defining the TTE endpoint.

7.5.1.2 Main Analytical Approach

For the analysis of primary efficacy endpoint, the time to event data up to Week 106 for SHP611-201 Group A and the GLIA-MLD matched external control group will be compared using the stratified log-rank test, where the matching identification created from the matching process in the SAS PSMATCH Procedure will be used as strata. The null and 1-sided alternative hypotheses of the stratified log-rank test are, respectively:

$$H_0: S_{\text{SHP611-201 Group A}} = S_{\text{GLIA-MLD matched external control}}$$

vs.

$$H_1: S_{\text{SHP611-201 Group A}} > S_{\text{GLIA-MLD matched external control}}$$

where $S_i = P(T_i > t)$ is the survival function in comparator arm $i \in$ {SHP611-201 Group A, GLIA-MLD matched external control group}, with T_i the corresponding survival time, a nonnegative random variable, and $t \geq 0$.

If the null hypothesis is rejected at the 1-sided 0.025 level, in the next step of the two-stage procedure (Rufibach 2019), the treatment effect will be quantified using a weighted average of proportion of patients not reaching the event of interest, with weights derived from the relative size of treated and control units in the strata used for the stratified log-rank test in the primary analysis. Interval censoring methods will be used, with event assumed to have first happened between the last visit/encounter prior to the event observation, and the visit/encounter when the event is first observed. Kaplan-Meier survival curves suitable for interval censoring data will be presented. Non-parametric estimates of the survival probability of not experiencing the event in the appropriate intervals (for both GLIA-MLD and SHP611-201), the treatment difference in survival probability between GLIA-MLD and SHP611-201 at Week 106 (or the last shared follow-up time), along with the corresponding 2-sided 95% confidence intervals, will be presented. See SAS sample code and description in Appendix 10.4 for more details, and the derivation of the respective survival probability standard errors for GLIA-MLD and SHP611-201. For the quantification of uncertainty in the treatment difference in survival probability between GLIA-MLD and SHP611-201, the variance of this difference will be calculated as the sum of the respective variances from GLIA-MLD and SHP611-201.

The age at the TTE starting point will be used to define day 0. Since the interval censoring method will be used, the age collected at the last visit/encounter prior to the event, and the age collected at the visit/encounter when the event is first observed, will be used to derive the interval for TTE endpoint.

In SHP611-201, Group A subjects who completed the primary treatment period but did not experience an event at or prior to the Week 106 visit (i.e., during the primary treatment period)

will be right censored at the actual Week 106 visit date, even if it falls beyond the exact date of Week 106. If Group A subjects discontinued early from the study during the primary treatment period without an observed event, subjects will be right censored at the last GMFC-MLD assessment time point (i.e., GMFC-MLD category should be no more than 4). If a subject had an event between the Week 79 visit date and Week 106 visit date, then the event will be considered to be within the interval of the two actual dates, even if the Week 106 visit date falls beyond the exact date of Week 106.

In GLIA-MLD, subjects who completed two years of follow up but did not experience an event at or prior to the first encounter after two years will be right censored at the first encounter after two years. If a subject had an event between two encounters that straddle across the 2-year time point (i.e., GLIA-MLD subjects who do not have an event at the last encounter before 2 years but did have an event observed at the first post 2-year encounter), the post 2-year encounter will be used to obtain the right end of the interval even if it is after 2 years.

7.5.1.3 Sensitivity Analysis

Sensitivity analysis 1:

This sensitivity analysis will be performed to compare the Study SHP611-201 Group A subjects vs. the GLIA-MLD matched external control group, but using the age at the Screening Visit as the TTE starting point for Study SHP611-201.

Sensitivity analysis 2:

The following sensitivity analyses may be conducted with alternative TTE starting point for the external control group from GLIA-MLD, if deemed necessary:

a. Using the first qualifying encounter as the TTE starting point
b. Using the last qualifying encounter as the TTE starting point
c. Including all qualifying encounters at category M0b, and treating indeterminate M0b's (defined below) as M1 when selecting the TTE starting point
d. Including all qualifying encounters at category M0b and treating indeterminate M0b's as M2 when selecting the TTE starting point
e. Excluding all qualifying encounters at category M0b when selecting the TTE starting point.

Note: In all analyses, the imputation of the M0b category will maintain the non-decreasing order across encounters. An "indeterminate M0b" is one where either a category M1 or M2 can be imputed while maintaining the order. In the following 3 examples, only the M0b in the first example is indeterminate. Only M2 can be imputed for the second example, and only M1 can be imputed for the third.

i) M1, M0b, M2
ii) M2, M0b, M3
iii) M0b, M1

Sensitivity analysis 3:

The following sensitivity analyses may be conducted with alternative matching and/or weighting options, if deemed necessary:

a. Optimal full matching would be performed using the average treatment effect (ATE) weight, instead of the ATT weight
b. The estimated propensity score for treatment group status (calculated using the 2 continuous covariates of age at MLD symptom onset and duration from MLD symptom onset to TTE starting point) would be used in optimal full matching. Specifically, the matching process is comprised of optimal full matching, minimizing Mahalanobis distance (based on the estimated propensity score and also the 2 covariates of age at MLD symptom onset and duration from MLD symptom onset to TTE starting point), with exact matching on GMFC-MLD category (M1 or M2)
c. All control subjects would be used without pruning
d. The estimated propensity score (ePS) for treatment group status (calculated using the 3 covariates of GMFC-MLD category, age at MLD symptom onset, and duration from MLD symptom onset to TTE starting point) would alternatively be used as weights in order to potentially enhance balance on covariates, ahead of endpoint analysis. Overlap weights would be applied to subjects in the SHP611-201 Group A and the external control subjects for Group A from GLIA-MLD, with weights of 1− ePS and ePS respectively, where ePS is the estimated propensity score associated with receiving IT administration of SHP611
e. Entropy balance (Hainmueller 2012) would be performed in which the statistician provides balancing criteria which are then employed to re-weight the control arm covariates to reflect those of the treatment, in turn keeping the treatment arm un-weighted

Sensitivity analysis 4:

This sensitivity analysis will be performed to estimate the unstratified survival probabilities. Specifically, the non-parametric estimates of the survival probability of not experiencing the event in the appropriate intervals (for both GLIA-MLD and SHP611-201), the treatment difference in survival probability between GLIA-MLD and SHP611-201 at Week 106 (or the last shared follow-up time), along with the corresponding 2-sided 95% confidence intervals, will be calculated based on the dataset prior to the matching process.

Sensitivity analysis 5:

This sensitivity analysis will be performed to compare the Study SHP611-201 Group A vs. the GLIA-MLD matched external control group, through a Cox proportional hazard model, i.e., a Cox proportional hazard model for interval censored data, based on the dataset prior to matching, will be applied using the following 3 covariates: age at MLD symptom onset, GMFC-MLD category at TTE starting point/Screening, and duration from MLD symptom onset to TTE starting point.

Sensitivity analysis 6:

The primary efficacy analysis will be repeated by using only the local GMFC-MLD assessments (both Screening and post-Screening assessments). The local GMFC-MLD category at Screening will also be used to re-derive the SHP611-201 Groups A assignment (Section 2.1), select the baseline encounter in GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3).

Sensitivity analysis 7:

As death information may not be collected for all of the matched external control subjects for Group A in GLIA-MLD, the primary efficacy analysis may be repeated, if deemed necessary, without treating death as an event (i.e. the event will only be defined by a progression to GMFC-MLD category 5 or higher, up to Week 106).

Sensitivity analysis 8:

The primary efficacy analysis will be repeated by including an additional filtering criterion (i.e., filtering criterion 4, in addition to the 3 filtering criteria in Section 3.1): "excluding GLIA-MLD untreated patients whose age at symptom onset was younger than the minimum age at symptom onset for the treated subjects in SHP611-201", to select external control subjects from the GLIA-MLD database.

Sensitivity analysis 9:

As study drug dosing was affected by the COVID-19 pandemic, a sensitivity analysis will be conducted for Group A subjects who missed no more than two consecutive SHP611 doses due to COVID-19.

Sensitivity analysis 10:

The primary efficacy analysis will be repeated by using the original adjudicated GMFC-MLD category at Screening for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 1.0. This original adjudicated GMFC-MLD category at Screening will be used to re-derive the SHP611-201 Groups A assignment (Section 2.1), select the baseline encounter in GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3). The centrally rated GMFC-MLD category (adjudicated as needed) for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 5.0, will be used for all post-Screening assessments.

Refer to Appendix 10.10 for a summary of the impact of the above sensitivity analyses on the primary efficacy analysis workflow.

7.5.2 Secondary Endpoints Analysis

1. Response in Group A, defined as maintenance of gross motor function at Week 106, evaluated as subjects who do not experience any event within Week 106, where event is defined as a decline in GMFC-MLD to category 5 or higher, or death

For this secondary efficacy endpoint, the event is defined as a decline in GMFC-MLD from the TTE starting point to category 5 or higher, or death, up to Week 106 (or 2 years), evaluated on subjects in SHP611-201 Group A or matched external control subjects. If a GMFC-MLD category M5b encounter is identified in a matched external control subject, M5b will be considered as part of the event definition for loss of locomotion (M5/M5b/M6 or death), therefore the encounter will be included in defining the binary endpoint.

For the secondary endpoint, the efficacy comparison will be assessed using the Cochran Mantel-Haenszel test at the 1-sided 0.025 level, where the matching identification created from the matching process in the SAS PSMATCH Procedure will be used as strata (with McNemar's test as a special case, if 1:1 matching is achieved). Confidence intervals of two-sided 95% coverage will be constructed for the proportion of subjects who do not experience any event within Week 106 for each comparator arm using the stratified Wilson confidence interval (Yan and Su 2010), and for the difference in proportions between the two comparator arms using the stratified Newcombe confidence interval, with Cochran-Mantel-Haenszel weights.

If a subject prematurely discontinues from the SHP611-201 study (the primary treatment period in SHP611-201), then "no maintenance of gross motor function" will be used as the outcome for the subject.

In the GLIA-MLD matched external control group, if subjects do not have complete information collected within the entire 2 years (such as a GMFC-MLD at category 4 or lower at the last encounter occurred prior to year 2), then "no maintenance of gross motor function" at 2 years will be used as outcome for the subject unless it could be supported by evidence that the subject still maintained gross motor function well with a GMFC-MLD category M4 or lower at certain encounter after 2 years.

This secondary efficacy endpoint analysis will be repeated by using the original adjudicated GMFC-MLD category at Screening for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 1.0. This original adjudicated GMFC-MLD category at Screening will be used to re-derive the SHP611-201 Group A assignment (Section 2.1), select the baseline encounter in GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3). The centrally rated GMFC-MLD category (adjudicated as needed) for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 5.0, will be used for all post-Screening assessments.

2a) Change from baseline at Week 106 and EOS in gross motor function, using the GMFC-MLD

This endpoint will be derived for the SHP611-201 subjects only. For this secondary endpoint, GMFC-MLD will be analyzed as an ordered categorical variable. For each GMFC-MLD category at baseline, the number and percentage of subjects for every possible categorical change from baseline (e.g., no change, 1 level progression, 2 level progression etc.) at Week 106 and EOS will be presented. The baseline value is selected as the GMFC-MLD assessment evaluated at Screening visit in SHP611-201.

In addition, GMFC-MLD categories will be summarized in terms of number and percentage of subjects in each level at each assessment time point, including a missing category if applicable, by subject group. The GMFC-MLD data by subject will also be presented in a listing.

An empirical cumulative distribution function (CDF) plot will also be generated for the change from baseline in GMFC-MLD.

For the GLIA-MLD matched external control group, all available GMFC-MLD data will also be presented in a by-subject listing.

2b) Subjects with unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2-categories that has not reverted to a 2-category decline (or better) at Week 106, evaluated on subjects in Group A

This endpoint will be derived for SHP611-201 group and GLIA-MLD matched external control group. For the secondary endpoint, the efficacy comparison will be assessed using the Cochran Mantel-Haenszel test at the 1-sided 0.025 level, where the matching identification created from the matching process in the SAS PSMATCH Procedure will be used as strata (with McNemar's test as a special case, if 1:1 matching is achieved) at Week 106. Confidence intervals of two-sided 95% coverage will be constructed for the proportion of subjects who do not experience unreversed decline from baseline in GMFC-MLD of more than 2 categories within Week 106 for each comparator arm using the stratified Wilson confidence interval (Yan and Su 2010), and for the difference in proportions between the two comparator arms using the stratified Newcombe confidence interval, with Cochran-Mantel-Haenszel weights.

This secondary efficacy endpoint analysis will be repeated by using the original adjudicated GMFC-MLD category at Screening for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 1.0. This original adjudicated GMFC-MLD category at Screening will be used to re-derive the SHP611-201 Group A assignment (Section 2.1), select the baseline encounter in GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3). The centrally rated GMFC-MLD category (adjudicated as needed) for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 5.0, will be used for all post-Screening assessments.

2c) Time to unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2 categories that has not reverted to a 2-category decline (or better) as of the last recorded observation

For this TTE endpoint, the TTE starting point in Group A subjects of SHP611-201 is set at the visit where patients receive the first dose of SHP611 (Visit 0). For the control group, the derivation of TTE starting point in GLIA-MLD is described in Section 3.2.

The event of unreversed decline from baseline in GMFC-MLD of more than 2 categories, defined as any decline of more than 2 categories that has not reverted to a 2-category decline (or better) up to Week 106 (or 2 years), will be evaluated on subjects in Group A (or matched external control subjects).

The weighted average of proportion of patients not reaching the event of interest will be calculated, with weights derived from the relative size of treated and control units in the strata used for the stratified log-rank test in the primary analysis. Interval censoring methods will be used, with event assumed to have first happened between the last visit/encounter prior to the event observation, and the visit/encounter when the event is first observed. Kaplan-Meier survival curves suitable for interval censoring data will be presented. Non-parametric estimates of the survival probability of not experiencing the event in the appropriate intervals (for both GLIA-MLD and SHP611-201), the treatment difference in survival probability between GLIA-MLD and SHP611-201 at Week 106 (or the last shared follow-up time), along with the corresponding 2-sided 95% confidence intervals, will be presented. See SAS sample code and description in Appendix 10.4 for more details, and the derivation of the respective survival probability standard errors for GLIA-MLD and SHP611-201. For the quantification of uncertainty in the treatment difference in survival probability between GLIA-MLD and SHP611-201, the variance of this difference will be calculated as the sum of the respective variances of GLIA-MLD and SHP611-201.

The age at the TTE starting point will be used to define day 0. Since the interval censoring method will be used, the age collected at the last visit/encounter prior to the event, and the age collected at the visit/encounter when the event is first observed, will be used to derive the interval for TTE endpoint.

In SHP611-201, Group A subjects who completed the primary treatment period but did not experience an event at or prior to the Week 106 visit (i.e., during the primary treatment period) will be right censored at the actual Week 106 visit date, even if it falls beyond the exact date of Week 106. If Group A subjects discontinued early from the study during the primary treatment period without an observed event, subjects will be right censored at the last GMFC-MLD assessment time point. If a subject had an event between the Week 79 visit date and Week 106 visit date, then the event will be considered to be within the interval of the two actual dates, even if the Week 106 visit date falls beyond the exact date of Week 106.

In GLIA-MLD, subjects who completed two years of follow up but did not experience an event at or prior to the first encounter after two years will be right censored at the first encounter after two years. If a subject had an event between two encounters that straddle across the 2-year time point (i.e., GLIA-MLD subjects who do not have an event at the last encounter before 2 years but did have an event observed at the first post 2-year encounter), the post 2-year encounter will be used to obtain the right end of the interval even if it is after 2 years.

3. Change from baseline at Week 106 and EOS in CSF sulfatides levels

This endpoint will be derived for the SHP611-201 group only. The values of change from baseline at Week 106 and EOS in CSF sulfatides levels will be summarized descriptively. The baseline value for CSF sulfatides is the one evaluated at Baseline visit in SHP611-201.

An empirical cumulative distribution function (CDF) plot will also be generated for the change from baseline in CSF sulfatides levels.

4. Response in Group A, defined as maintenance of gross motor function at Week 106, defined as a GMFM-88 total score ≥40

GMFM-88 total score (percent): Calculated by averaging the percent scores programmatically for each of the 5 domains (i.e., lying and rolling; sitting; crawling and kneeling; standing; and walking, running, and jumping) and rounding to the nearest whole number based on the GMFM-88 scoring rule.

This binary endpoint will be derived for SHP611-201 subjects only. For this secondary endpoint, it will be analyzed by the counts and proportions of each value, including a missing category if applicable. The denominator for the proportion will be based on the number of subjects who provided non missing responses to the binary variable.

5a) Time to unreversed decline from baseline at Week 106 and EOS in GMFM-88 total score i.e. a decrease of >20 points or unreversed decline to a score <40 points, whichever occurs first

This endpoint will be derived for SHP611-201 subjects only. Kaplan-Meier survival curve will be presented.

This time to event endpoint will use GMFM-88 data collected up to Week 106. If subjects did not experience an event, which is defined as an unreversed decline from baseline in GMFM-88 total score of >20 points or unreversed decline to <40 points, whichever occurs first, subjects are censored at the last recorded observation.

Interval censoring method will be used for this endpoint, the age at the TTE starting point, the age collected at the last visit prior to the event, and the age collected at the visit when the event is first observed, will be used to derive the TTE endpoint.

5b) Change from baseline at Week 106 and EOS of gross motor function, using the GMFM-88 total score

This endpoint will be derived for SHP611-201 subjects only. For this secondary endpoint, the change from baseline at Week 106 and EOS in GMFM-88 will be summarized descriptively.

The GMFM-88 assessments (including the total score and score in each domain) at each assessment time point will also be presented in a by subject listing.

5c) Subjects in Group A with GMFM-88 total score decrease of ≤20 points from baseline and a total score that is ≥40 at Week 106 and EOS

This endpoint will be derived for SHP611-201 subjects only. For this secondary endpoint, subjects will be dichotomized according to whether their GMFM-88 total score decline is ≤20 points from baseline and a total score that is ≥40 at Week 106 and EOS, respectively.

This endpoint will be summarized descriptively.

6. Change from baseline at Week 106 and EOS in expressive language using the ELFC-MLD

ELFC-MLD is a categorical scale of 0-4 levels (Refer to Appendix 10.8).

This endpoint will be derived for SHP611-201 subjects only. Change from baseline at Week 106 and EOS in ELFC-MLD will be summarized descriptively among subjects with a value at both baseline and the specific post-baseline visit. The baseline value in ELFC-MLD assessment is the one evaluated at Baseline visit in SHP611-201.

7.5.3 Subgroup Analyses

For subjects in SHP611-201 Groups A, B, C, D, E, and F, descriptive summaries of the efficacy data will be provided.

Only exploratory statistical evaluations involving Groups B, C, D, E, and F may be conducted.

In addition, descriptive summaries of the primary endpoint efficacy data will be provided by

- sex
- race
- ethnicity and
- geographic region

If the number of subjects in the subgroup is less than 3, only a listing will be provided instead of the descriptive statistics.

7.6 Pharmacokinetic Analyses

All PK analyses will be performed using the PK Analysis Set.

Blood and CSF samples will be collected for determination of SHP611 levels after IT administration. SHP611 concentrations in serum and CSF will be determined using a validated Enzyme-Linked Immunosorbent Assay (ELISA) method which was used for the previous SHP611 clinical studies (HGT-MLD-070/HGT-MLD-071). SHP611 activity in CSF and serum will also be determined by a validated Activity assay. The SHP611 Activity results will be used as a surrogate marker for anti-SHP611 neutralizing antibodies and its impact on PK profile.

Details of the PK analysis including handling of PK data, parameters estimated, and presentation of PK data will be provided in the Clinical Pharmacology Analysis Plan (CPAP).

There will be no inferential statistical analysis of the PK data. Summary statistics (number of observations [N], mean, SD, coefficient of variation [CV%], median, maximum, minimum, geometric mean and geometric CV%) will be determined for all serum PK parameters and presented by bioanalytical method and visit for each group and for overall population. Serum and CSF concentrations at each nominal sampling time will also be summarized by bioanalytical method and visit for each group and for overall population using descriptive statistics. Any additional details will be provided in the CPAP.

7.7 Safety Analysis

This section will only apply to SHP611-201. All safety analyses will be based on the Safety Analysis Set.

7.7.1 Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as AEs that occurred at or after the first dose of investigational product or device implant surgery (whichever occurs first) and through the last follow-up date plus 14 days (inclusive). Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 21.0 or higher. The number of events and percentage of TEAEs will be calculated overall, by system organ class (SOC), by preferred term, by subject groups (A-F) and overall. TEAEs will be further summarized by severity, relationship to investigational product, disease and outcomes, the IDDD, the IDDD surgical procedure, anesthesia, and IT administration process. Adverse events related to investigational product, AEs leading to withdrawal, serious adverse events (SAEs; all SAEs reported by the investigator, those SAEs considered as Related and those considered as Not Related by Takeda shall be collected), and deaths will be similarly summarized and listed.

Incomplete AE start date imputation rules:

The following rules will be applied to impute the missing numerical fields. If the stop date is complete and the imputed start date is after the stop date, then the start date will be imputed using the stop date.

Missing day and month

- If the year of the incomplete start date is the same as the year of the date of the first dose of investigational product, then the day and month of the date of the first dose of investigational product will be assigned to the missing fields
- If the year of the incomplete start date is before the year of the date of the first dose of investigational product, then December 31 will be assigned to the missing fields
- If the year of the incomplete start date is after the year of the date of the first dose of investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete start date are the same as the month and year of the date of the first dose of investigational product, then the day of the date of the first dose of investigational product will be assigned to the missing day
- If either the year is before the year of the date of the first dose of investigational product or if both years are the same but the month is before the month of the date of the first dose of investigational product, then the last day of the month will be assigned to the missing day

- If either the year is after the year of the date of the first dose of investigational product or if both years are the same but the month is after the month of the date of the first dose of investigational product, then the first day of the month will be assigned to the missing day.

Incomplete AE stop date imputation rules:

The following rules will be applied to impute the missing numerical fields. If the date of the last dose of investigational product is missing, then replace it with the last visit date. If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be equal to the start date.

Missing day and month

- If the year of the incomplete stop date is the same as the year as of the date of the last dose of investigational product, then the day and month of the date of the last dose of investigational product will be assigned to the missing fields
- If the year of the incomplete stop date is before the year of the date of the last dose of investigational product, then 31 December will be assigned to the missing fields
- If the year of the incomplete stop date is after the year of the date of the last dose of investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete stop date are the same as the month and year of the date of the last dose of investigational product, then the day of the date of the last dose of investigational product will be assigned to the missing day
- If either the year is before the year of the date of the last dose of investigational product or if both years are the same but the month is before the month of the date of the last dose of investigational product, then the last day of the month will be assigned to the missing day
- If either the year is after the year of the last dose of investigational product or if both years are the same but the month is after the month of the date of the last dose of investigational product, then the first day of the month will be assigned to the missing day.

7.7.2 Other Safety Analysis

Clinical laboratory tests, vital signs, and ECG findings will be summarized by subject group and visit. Potentially clinically important findings will also be summarized and listed. Descriptive summaries will also be provided for 12-lead ECG, CSF laboratory parameters (chemistries, cell counts), anti-SHP611 antibodies in CSF and serum, and determination of antibodies having enzyme neutralizing activity.

7.7.2.1 Clinical and CSF Laboratory Evaluations

The laboratory tests include hematology, clinical chemistry, urinalysis and cerebrospinal fluid (CSF) measurements (Biomarkers). All laboratory summaries will be performed using the Safety Analysis Set.

Descriptive statistics for clinical and CSF laboratory values and changes from baseline at each visit will be presented by subject group for the following clinical and CSF laboratory variables.

Hematology	Hemoglobin, hematocrit (HCT), red blood cells (RBC), platelet count, white blood cell (WBC) count – total and differential.
Chemistry	Albumin, Alkaline phosphatase (ALP), Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Amylase, Blood urea nitrogen (BUN), Calcium. Creatinine, Creatine kinase, Gamma-glutamyl transferase (GGT), Inorganic phosphate, Iron, Lactate dehydrogenase (LDH), Magnesium, Potassium, Sodium, Total bilirubin (BILI)
Urinalysis	Glucose, specific gravity, ketones, protein, bilirubin, pH, nitrite.
Biomarkers	Serum sulfatide, CSF sulfatide, CSF lyso-sulfatide, Normalized Urine sulfatide
CSF	Albumin, cell count, protein, glucose, Lymphocytes/Leukocytes differential, Monocytes/Leukocytes differential

Additionally, shift tables providing the number of subjects with indicated shifts (low, normal, high) in their results from Baseline to all available post baseline visits will also be presented for all parameters within Hematology, Chemistry and Urinalysis.

All laboratory data will be listed for the Safety Analysis Set. Change from Baseline over time plots will be provided.

7.7.2.2 Vital Signs

All vital signs summaries will be performed using the Safety Analysis Set.

Vital signs will include measurements of blood pressure, heart rate, respiratory rate, and body temperature. Vital signs will be measured within 30 minutes prior to IT administration of SHP611 and 30 (±5), minutes, 60 (±5) minutes, and 120 (±5) minutes post IT SHP611 administration. Height, weight, and head circumference will be recorded as part of vital signs, but will be done less frequently than vital signs per the Schedule of Activities from protocol.

Descriptive statistics for vital signs and their changes from baseline at each assessment time point will be summarized for each subject group and overall.

All vital sign data by subject will be presented in a listing. Change from Baseline over time plots will be provided.

7.7.2.3 Physical examination

If results of the physical examination show clinically significant worsening from the previous visit, the change will be documented as an AE/serious AE (SAE) in the eCRF.

Abnormalities identified at the Screening Visit will be documented in the subject's source documents and on the medical history eCRF. Changes after the Screening Visit will be captured as AEs on the AE eCRF page, as deemed appropriate by the investigator.

Refer to Section 7.7.1 for AE analysis.

7.7.2.4 Electrocardiogram (ECG)

ECG measurements will include heart rate, RR, PR, QRS and QT duration. Descriptive statistics for ECG variables and their changes from baseline at each assessment time point will be presented by subject group. All ECG data by subject will be presented in a listing. Change from Baseline over time plots will be provided.

7.7.2.5 CSF laboratory parameters

Descriptive statistics for CSF laboratory parameters (chemistries, cell counts) and their changes from baseline at Week 106 and EOS will be presented by subject group. All CSF laboratory parameters data by subject will be presented in a listing. Change from Baseline over time plots will be provided.

7.7.2.6 IDDD Performance

SOPH-A-PORT® safety and performance will be summarized for IDDD-implanted subjects in Safety Analysis Set. The number and proportion of subjects of the following categories and the corresponding event count and event percentage will be summarized.

1. with the initial implant only (i.e., no additional surgeries)
2. who had any post-initial implantation device surgeries
3. who had difficulties associated with the implant procedure (e.g., Difficulty Accessing Spinal Canal, etc)

The possible outcomes of a device malfunction are resolved, ongoing or failure. If the outcome of a malfunction is missing, it is assumed to be ongoing for purposes of analysis (i.e. a malfunction, but not yet resolved or a failure). The number and proportion of subjects and IDDDs with one or more: total malfunctions (including malfunctions with an outcome of failure, resolved or on-going); malfunctions leading to failure; resolved malfunctions; and on-going malfunctions; as well as the corresponding event numbers will be presented. The types of total malfunctions and the reasons for IDDD failures reported by the site will be summarized at the subject, IDDD and event level. A by-subject listing of the device failure and resolved malfunction data will be displayed.

The overall IDDD failure rate and its 95% CI of subjects will be presented within one table. The overall IDDD failure rate is calculated as the total number of IDDD failures for all subjects divided by the total IDDD time in years at risk, which is defined as the total time to IDDD failure

or the last injection if IDDD is not failed at the time of the data cut, from initial implantation, delayed implantation, or revision for all IDDDs.

The IDDD longevity (time to failure in weeks), for all implanted IDDDs, will be plotted using the Kaplan Meier (KM) method. A new port identifies a new IDDD, starting from the date of implantation (either an initial implantation, a partial or full revision, or a delayed device implant after previous removal). The time to IDDD failure (weeks) will be obtained by subtracting the date of the IDDD kit implantation from the date of IDDD failure (i.e., the initial malfunction date that persists leading to surgical intervention) plus 1, and divided by 7, one decimal will be kept. IDDDs which did not fail will be censored at the last drug injection date on or prior to each data cut for each IDDD.

An IDDD Timeline of events (in weeks from implantation) including the timing of device surgeries (adjustment, partial or full revision, Complete Device Removal without Immediate Replacement, and Delayed Device Implant after Previous Removal), and malfunction and failure by IDDDs will be plotted by subject sorted by the duration since the initial implantation, longest on top.

The number of surgeries per subject will be presented by a bar chart. The different types of surgeries (All Surgeries including Initial Implant, Device Adjustment, Removal and Replacement of Port Only, Complete Removal and Replacement, Complete Device Removal Only, Delayed Device Implant after Previous Removal, and Other) will be presented as different colors and/or patterns within the same bar.

7.7.3 Extent of Exposure and Compliance

The total number of doses of study drug, the number of doses received via IDDD, the number of doses received via lumbar puncture (LP), and treatment compliance will be summarized by subject group for the Safety Analysis Set.

Treatment compliance is defined as: [(Number of Complete IT administrations) ÷ (Expected Number of IT administrations at the time of each data cut)] * 100.

Exposure to study drug for the Safety Analysis Set will be summarized in terms of treatment duration, which is calculated as the number of days from the date of first dose of investigational product taken to the date of the last dose of investigational product taken, inclusively.

Duration of exposure is defined as: date of the last dose of investigational product taken – date of first dose of investigational product taken + 1

Descriptive statistics will be presented to describe the exposure to investigational product by subject group and overall.

7.8 Immunogenicity Analysis

All immunogenicity-related analyses will be performed using the Immunogenicity Analysis Set.

Serum and CSF samples will be collected at scheduled visits as described in Schedule of Activities per protocol (see Section 2.1) for monitoring the formation of anti-SHP611 antibodies (ADA) throughout the study.

Based on the availability of immunogenicity data, as applicable, the following analyses will be performed to characterize the incidence, onset and duration of ADA including neutralizing antibodies, titers, and the impact of ADA on PK, PD, efficacy and safety:

- The number and percentage of patients with pre-existing (baseline), treatment-induced, treatment-boosted and overall incidences of ADA in serum and CSF will be summarized by visit for each group and for overall patients. Similar summary will be provided for serum neutralizing antibodies.
- A boxplot and descriptive summary of serum and CSF ADA titer values by visit for each group and for overall patients.
- A boxplot and descriptive summary of SHP611 CSF C_{trough} values in ADA-positive and ADA-negative patients by visit.
- A boxplot and descriptive summary of SHP611 serum C_{trough} values in ADA-positive and ADA-negative patients by visit.
- A boxplot and descriptive summary of SHP611 serum CL/F values in ADA-positive and ADA-negative patients by visit.
- Summary and plot of key efficacy and PD endpoints in CSF ADA-positive and ADA-negative patients over time.

A by-subject listing of antibody results will be presented.

A subject is considered to be ADA positive at Week 106 if there is at least 1 ADA positive result in the primary treatment period. Transiently positive ADA is defined as subjects who have confirmed positive ADA status at 1 time point during treatment or follow-up period (excluding the last sampling time point) or at 2 or more time points during treatment where first and last positive samples are separated by less than 16 weeks with last sampling time point testing negative. Persistently positive ADA is defined as subjects who have confirmed positive ADA status at 2 or more time points where first and last positive samples are separated by 16 weeks or longer, or ADA positive only at the last sampling time point or within 16 weeks to the last time point that is negative. Treatment-induced ADA is defined as subjects with baseline negative ADA and at least 1 post-baseline ADA positive status. Treatment-boosted ADA is defined as subjects with baseline positive ADA and at least 1 post-baseline ADA positive with titer >4 folds of baseline titer value.

7.9 Exploratory Analyses

Exploratory statistical evaluations will be performed for all subjects from Groups A, B, C, D, E, and F in SHP611-201.

Descriptive analyses and/or listings, as deemed appropriate, will be provided for the following exploratory endpoints. See Section 7.1 for general analysis approach for each type of endpoint.

1. Change in CSF, serum, and urine biomarkers over time

Observed values for sulfatide and lyso-sulfatide will be summarized at Baseline and at all post baseline visits by subject group and overall. The change from baseline and percent change from baseline by visit will also be summarized by subject group and overall.

Descriptive statistics will be presented. Furthermore, corresponding mean plots with SE bars and boxplots for each subject group will be plotted at each appropriate visit, and data for individual subjects will also be listed by subject group and visit.

2. Change in MRS metabolite levels specifically: N-acetylaspartate/Creatine over time

The values of change from baseline at Week 106 and EOS in MRS metabolite levels will be summarized descriptively. The baseline is the assessment evaluated prior to first IT administration of SHP611.

3. Change in Eichler MLD MRI severity score over time

Change from baseline at Week 106 and EOS in Eichler MLD MRI severity score will be summarized descriptively. The baseline is the assessment evaluated prior to first IT administration of SHP611.

4. Change in total MLD severity score based on brain MRI over time

Based on a visual scoring method of the MRI, a total MLD severity score will be calculated for each subject where higher scores indicate more severe brain involvement. The observed values at each visit as well as the change from baseline to all post-baseline visit in the total MLD severity score will be summarized by subject group and overall.

5. Change in volumetric analysis of the brain based on MRI over time

The values of the change from baseline at Week 106 and EOS in volumetric analysis of the brain based on MRI will be summarized by subject group and overall.

6. Change in GIMF-C over time

The GIMF-C will be rated based on the clinician's assessment of any changes in a patient's motor function since the baseline study visit (Visit 0) on a 5-point scale ranging from "much improved" to "much worse". It measures the changes in overall gross motor function and in each of the five dimensions of the GMFM-88 (Lying and rolling, Sitting, Crawling and kneeling, Standing and Walking, running and jumping).

Descriptive statistics for each dimension of GIMF-C at each post-baseline visit will be presented by subject group. The GIMF-C by subject will also be presented in a listing.

7. Change in GIMF-S over time

The GIMF-S will be rated based on the clinician's assessment of a patient's current gross function severity on a 5-point scale ranging from "not impaired at all" to "totally impaired (no gross motor function)". It measures the current impairment level in overall gross motor function and in each of the five dimensions of the GMFM-88 (Lying and rolling, Sitting, Crawling and kneeling, Standing and Walking, running and jumping).

Descriptive statistics for each dimension of GIMF-S at each visit will be presented by subject group and overall. Changes from baseline analysis will be conducted and will include only subjects with a value at both baseline and the specific post-baseline visit. The GIMF-S by subject will also be presented in a listing.

8a). Descriptive statistics of the Caregiver Impact Questionnaire (CIQ) item responses over time to inform scoring

Caregiver burden will be assessed by the Caregiver Impact Questionnaire (CIQ). The CIQ includes 30 items in total and covers the key areas of impact for caregivers of patients with late infantile and juvenile MLD, including: 1) impact on relationships, family, social life, and leisure activities; 2) impact on personal time and daily activities; 3) emotional/ psychological impacts; 4) impact on physical health; and 5) impact on finances and productivity. The individual item responses are scored using the following response scale: 0 = never, 1 = rarely, 2 = sometimes, 3 = often, and 4 = always, with higher scores corresponding to higher degree of impact on caregivers. Scores for each domain will be calculated as the average of the items in that domain (Social functioning – 7 items; Impact on daily activities – 5 items; Emotional functioning – 10 items; Psychological functioning – 6 items; Financial impact – 2 items). CIQ total score is calculated as the average of the completed domain scores.

Descriptive statistics for CIQ domain at each visit will be presented by subject group. Changes from baseline analysis will be conducted and will include only subjects with a value at both baseline and the specific post-baseline visit. The CIQ by subject will also be presented in a listing.

8b). Change in each of the parent and infant/toddler concepts as assessed by the Infant Toddler Quality of Life Questionnaire – 97 items (ITQOL-97) over time

The ITQOL-97 was developed for use in infants and toddlers at least 2 months of age up to 5 years measuring the World Health Organizations definition of health as a state of complete physical, mental and social well-being. It measures infant/toddler focused concepts and parent-focused concepts. Infant/toddler concepts include overall health (1 item), amount of limitation in physical activities (10 items), satisfaction with development (10 items), amount, frequency of bodily discomfort and the extent to which pain/discomfort interferes with normal activities (3 items), frequency of certain moods and temperaments (18 items), perceptions of current, past and future behavior (12 items), overall behavior (1 item) and frequency of behavior problems (15 items), perception of current, past and future health (11 items), perceptions of changes in health over the past year (1 item). Parent focused concepts include amount of worry experienced by parent (7 items), amount of time limitations experienced by parent (7 items) and rating time of family's ability to get along with one another (1 item). For each concept, item responses will be scored, summed, and transformed to a scale from 0 (worst health) to 100 (best health) according to the scoring rules from the ITQOL-97 licenser.

Descriptive statistics for each of the parent and infant/toddler concepts at each visit will be presented by subject group and overall. Changes from baseline analysis will be conducted and will include only subjects with a value at both baseline and the specific post-baseline visit.

9. Change in incidence of hospitalizations, number of days in hospital, reason for admission, and frequency of selected MLD-related procedures (use of feeding tube, use of intubation, type of respiratory support) over time; total number of additional hospitalizations during the 2-year follow-up

Healthcare Utilization Questionnaire (HCUQ) variables include the number of hospitalizations, days in the hospital, and reasons for admission, as well as use of selected MLD-related procedures including use of a feeding tube, intubation, and type of respiratory support. Descriptive statistics for these key HCUQ variables at each visit will be presented.

10. Change in work productivity and activity impairment as assessed using the Work Productivity and Activity Impairment Questionnaire (WPAI): Specific Health Problem V2.0 over time

The instrument has 6 questions. The scoring is based on the scoring rules on the licensor's website (http://www.reillyassociates.net/WPAI_Scoring.html) and four metrics of WPAI measurement will be calculated:

1. Absenteeism (work time missed)
2. Presenteeism (impairment at work / reduced on-the-job effectiveness)
3. Work productivity loss (overall work impairment / absenteeism plus presenteeism)
4. Activity Impairment

Descriptive statistics for each of the WPAI measurements at each visit will be presented by subject group and overall. Changes from baseline analysis will be conducted and will include only subjects with a value at both baseline and the specific post-baseline visit.

11. Change in ability to eat and drink as assessed using the Eating and Drinking Ability Classification System (EDACS) assessments over time

EDACS is a categorical scale of 1-5 levels (Refer to Appendix 10.9). Descriptive statistics at each available visit will be presented by subject group and overall. Changes from baseline analysis will be conducted and will include only subjects with a value at both baseline and the specific post-baseline visit.

12. Restricted mean survival time, defined as the average time free from an event up to Week 106, where event is defined as a decline in GMFC-MLD to category 5 or higher, or death

This analysis will be performed to compare the Study SHP611-201 Group A vs. the GLIA-MLD matched external control group in restricted mean survival time, i.e., the treatment difference (with 2-sided 95% confidence interval) in the restricted mean survival time for interval-censored data (Zhang et al. 2020) by Week 106 will be estimated. This analysis will be performed based on two datasets: (a) dataset prior to the matching process and (b) weighted dataset after the matching process.

7.10 Extension Period

Summary statistics for continuous variables will include the number of subjects, mean, standard deviation (SD), median, minimum, and maximum. Categorical variables will be summarized using the number and percentage of subjects in each category, including a missing category if applicable.

Statistical inferences for the primary and selected secondary endpoints may be performed during the extension period.

The safety analyses will be performed using all data through Week 106 and the extension period up to EOS. Baseline will be defined as the same as that for the study (this baseline is clinically relevant for assessing the long-term safety and efficacy outcomes of extended treatment with IT SHP611).

The same analysis populations, as in the primary treatment period, will be used for the extension period.

Device-related analyses will be conducted in the subset of subjects in the Safety Analysis Set who had the device implant procedure performed.

7.11 Interim Analyses

No formal interim analysis is planned for this study.

7.12 Data Monitoring Committee

An external independent Data Monitoring Committee (DMC) will be involved in the conduct of this study. The purpose of the DMC is to review the data pertaining to safety, tolerability, and benefit/harm of the study therapy for the duration of the study. The DMC will oversee both administration of IT SHP611 and device safety. The DMC will be notified of IDDD failures and related complications on a periodic basis. Further details regarding the DMC can be found in the DMC Charter, which will be available prior to the administration of investigational product.

8. REFERENCES

Austin, P. C. and Stuart, E. A. 2015. Optimal full matching for survival outcomes: a method that merits more widespread use. *Stat Med,* 34(30), 3949-67.

Hainmueller, J. 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit Anal,* 20(1), 25-46.

Kehrer, C., Blumenstock, G., Gieselmann, V. and Krageloh-Mann, I. 2011b. The natural course of gross motor deterioration in metachromatic leukodystrophy. *Dev Med Child Neurol,* 53(9), 850-5.

Kehrer, C., Blumenstock, G., Raabe, C. and Krageloh-Mann, I. 2011a. Development and reliability of a classification system for gross motor function in children with metachromatic leucodystrophy. *Dev Med Child Neurol,* 53(2), 156-60.

Rufibach, K. 2019. Treatment effect quantification for time-to-event endpoints-Estimands, analysis strategies, and beyond. *Pharm Stat,* 18(2), 145-65.

Yan, X. and Su, X. G. 2010. Stratified Wilson and Newcombe confidence intervals for multiple binomial proportions. *Stat Biopharm Res,* 2(3), 329-35.

Zhang, C., Wu, Y. and Yin, G. 2020. Restricted mean survival time for interval-censored data. *Stat Med,* 39(26), 3879-95.

9. CHANGES TO PROTOCOL PLANNED ANALYSES

No changes from protocol specified analyses are planned.

10. APPENDIX

10.1 Changes From the Previous Version of the SAP

This is Version 2.0 of the SAP. Changes from Version 1.0 of the SAP are detailed below.

SAP Section	Impacted Text (shown in bold)	Summary of Change	Rationale for Change
3.3.1 (Matching Diagnostics)	**Weighted** plots will be presented to assess variable differences between the SHP611-201 group and the matched external control group for all variables, **with weights derived from the respective matching or weighting process**	Weights derived from the respective matching or weighting process will be applied to the matching diagnostics plots	Weighted matching diagnostics plots are needed to evaluate the degree of cohort balance achieved through matching or weighting
7.1 (General Considerations)	**The centrally rated GMFC-MLD category (adjudicated as needed) for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 5.0, will be used for all corresponding inferential and descriptive analyses. This centrally rated GMFC-MLD category at Screening will also be used to define SHP611-201 subject groups A-F (Section 2.1), select the baseline encounter in GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3).**	Pre-specify the use of centrally rated GMFC-MLD category for all primary analyses	There are 3 types of GMFC-MLD assessments (central, local, and original adjudicated score at Screening). The centrally rated GMFC-MLD category will used for all primary analyses
7.5.1.2 (Main Analytical Approach) and 7.5.2 (Secondary Endpoints Analysis 2c)	**See SAS sample code and description in Appendix 10.4 for more details, and the derivation of the respective survival probability standard errors for GLIA-MLD and SHP611-201. For the quantification of uncertainty in the treatment difference in survival probability between GLIA-MLD and SHP611-201, the variance of this difference will be calculated as the sum of the respective variances of GLIA-MLD and SHP611-201**	Addition of details for the quantification of uncertainty in survival probability estimation	Additional details for the quantification of uncertainty are provided for the calculation of survival probability standard error and the corresponding 95% CI

SAP Section	Impacted Text (shown in bold)	Summary of Change	Rationale for Change
7.5.1.3 (Sensitivity Analysis 3)	3b. The estimated propensity score for treatment group status (calculated using the 2 continuous covariates of age at MLD symptom onset and duration from MLD symptom onset to TTE starting point) would be used in optimal full matching. Specifically, **the matching process is comprised of optimal full matching, minimizing Mahalanobis distance (based on the estimated propensity score and also the 2 covariates of age at MLD symptom onset and duration from MLD symptom onset to TTE starting point), with exact matching on GMFC-MLD category (M1 or M2)** 3d. **The estimated propensity score (ePS) for treatment group status (calculated using the 3 covariates of GMFC-MLD category, age at MLD symptom onset, and duration from MLD symptom onset to TTE starting point)** would alternatively be used as weights in order to potentially enhance balance on covariates, ahead of endpoint analysis. Overlap weights would be applied to subjects in the SHP611-201 Group A and the external control subjects for Group A from GLIA-MLD, with weights of 1− ePS and ePS respectively, where ePS is the estimated propensity score associated with receiving IT administration of SHP611	Addition of details for the sensitivity analyses on propensity score matching and weighting, and deletion of the cardinality matching sensitivity analysis	More details are provided for appropriate alternative matching and weighting approaches in the sensitivity analysis section
7.5.1.3 (Sensitivity Analysis 6) and 7.9 (Exploratory Analysis 12)	**The primary efficacy analysis will be repeated by using only the local GMFC-MLD assessments (both Screening and post-Screening assessments). The local GMFC-MLD category at Screening will also be used to re-derive the SHP611-201 Groups A assignment (Section 2.1), select the baseline encounter in**	Pre-specify the use of locally rated GMFC-MLD category for the sensitivity analysis. The original sensitivity analysis 6 on restricted mean survival time was moved to the exploratory analysis section	There are 3 types of GMFC-MLD assessments (central, local, and original adjudicated score at Screening). The locally rated GMFC-MLD category will be used for the sensitivity analysis

SAP Section	Impacted Text (shown in bold)	Summary of Change	Rationale for Change
	GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3)		
7.5.1.3 (Sensitivity Analysis 10) and 7.5.2 (Secondary Endpoints Analyses 1 and 2b)	**The primary efficacy analysis will be repeated by using the original adjudicated GMFC-MLD category at Screening for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 1.0. This original adjudicated GMFC-MLD category at Screening will be used to re-derive the SHP611-201 Groups A assignment (Section 2.1), select the baseline encounter in GLIA-MLD (Section 3.2), and conduct optimal full matching (Section 3.3). The centrally rated GMFC-MLD category (adjudicated as needed) for study SHP611-201, based on the SHP611-201 Video Acquisition & Adjudication Charter Version 5.0, will be used for all post-Screening assessments**	Pre-specify the use of original adjudicated GMFC-MLD category at Screening for the sensitivity analyses (for primary and 2 secondary efficacy endpoints)	There are 3 types of GMFC-MLD assessments (central, local, and original adjudicated score at Screening). The original adjudicated GMFC-MLD category at Screening will be used for the sensitivity analyses
10.2 (Definition of Baseline)	For SHP611-201 subjects, the respective GMFC-MLD and **GMFM-88** assessments at the Screening visit will be treated as the last clinically valid assessment prior to the administration of the investigational product, provided that the first dose of SHP611 is administered within 28 days of the GMFC-MLD assessment at Screening, per protocol. In the event that the first dose of SHP611 is administered more than 28 days after the Screening GMFC-MLD assessment, the respective GMFC-MLD and **GMFM-88** assessments performed on the day of administration of the first dose of SHP611 (Visit 0), per protocol, will be used as Baseline	Addition of Baseline definition for GMFM-88	Addition of Baseline definition for GMFM-88, to be consistent with the protocol schedule of assessment

10.2 Definition of Baseline

In general, Baseline for the treatment group is defined as the last assessment prior to the first administration of the investigational product, unless other clarification is made. For SHP611-201 subjects, the respective GMFC-MLD and GMFM-88 assessments at the Screening visit will be treated as the last clinically valid assessment prior to the administration of the investigational product, provided that the first dose of SHP611 is administered within 28 days of the GMFC-MLD assessment at Screening, per protocol. In the event that the first dose of SHP611 is administered more than 28 days after the Screening GMFC-MLD assessment, the respective GMFC-MLD and GMFM-88 assessments performed on the day of administration of the first dose of SHP611 (Visit 0), per protocol, will be used as Baseline. For the external control subjects for Group A from GLIA-MLD, the derivation of the baseline encounter (TTE starting point) is described in Section 3.2.

10.3 List of Terms for Filtering Criteria 2 for GLIA-MLD, 'Documented Gait Disorder'

Spasticity of extremities

Truncal hypotonia

Gait/truncal ataxia

Appendicular ataxia

Dystonia

Rigidity

Choreoathetosis

Movement Abnormalities

Loss of Assisted Ambulation

Loss of Independent Ambulation

Not able to get to sitting

Not able to place hands together

Not able to pull to stand

Not able to reach for objects

Not able to sit with no support

Not able to smile responsively

Not able to stand alone

Not able to use thumb-finger grasp

Not able to walk up steps

Not able to walk well

Not able to stand while holding on

Head control lost

Not able to roll over

10.4 SAS Sample Code for Optimal Full Matching Procedure and Primary Efficacy Analysis

```
/********** 1. Sample Data (dummy data) *************/

/*
Variables:
   Study: SHP611 vs GLIA-MLD
   PatientID: Id of the patient
   BaseGMFC: Baseline GMFC Score (1 or 2)
   OnsetAge: Age at MLD symptom onset
   BaseAge: Age at Index Date
   Duration: BaseAge - OnsetAge
   lt: Lower Limit of the Interval (after this encounter/visit Endpoint
was observed)
   rt: Upper Limit of the Interval (before this encounter/visit
Endpoint was observed)
*/

data sample;
input Study $ PatientID BaseGMFC OnsetAge BaseAge lt rt;
Duration = BaseAge-OnsetAge;
datalines;
GLIA-MLD 1  1 20.2341 31.3650 0.00 1.00
GLIA-MLD 2  2 15.2988 18.0298 0.00 1.00
GLIA-MLD 3  2 18.5828 29.4586 0.10 0.30
GLIA-MLD 4  1 24.4644 25.0197 0.90 .
GLIA-MLD 5  2 26.3237 27.3234 0.80 1.00
GLIA-MLD 6  2 24.9761 25.5248 0.10 0.30
GLIA-MLD 7  2 20.6306 27.0216 1.00 .
GLIA-MLD 8  2 15.9073 28.8836 0.00 0.10
GLIA-MLD 9  2 21.0788 23.8088 0.00 1.00
GLIA-MLD 10 1 13.3083 24.6536 0.60 .
GLIA-MLD 11 2 15.4416 27.3955 0.60 0.90
GLIA-MLD 12 1 21.3905 23.5588 0.20 0.40
GLIA-MLD 13 1 24.2239 25.0094 0.00 0.30
GLIA-MLD 14 2 22.2359 22.7278 0.10 0.30
GLIA-MLD 15 2 21.8668 22.3588 0.10 .
GLIA-MLD 16 2 24.8161 30.2214 0.00 1.00
SHP611  1  2  7.4770 33.7193 2.04 .
SHP611  2  2 20.0011 33.6061 2.04 .
SHP611  3  2 13.3568 38.0798 0.50 0.77
SHP611  4  1  9.3548 24.8425 2.04 .
SHP611  5  1 18.2042 30.4110 2.04 .
SHP611  6  2  7.4743 25.4382 2.04 .
SHP611  7  1 11.6451 35.1017 0.50 0.77
SHP611  8  2 26.4492 27.2987 2.04 .
SHP611  9  2  7.1980 38.8606 2.04 .
SHP611  10 1 29.8740 35.6204 2.04 .
```

```
SHP611  11 2  9.1920 26.3609 2.04 .
SHP611  12 2 17.5906 24.9879 0.00 0.10
SHP611  13 1  2.3253 36.2661 2.04 .
SHP611  14 1 28.9509 35.0880 2.04 .
SHP611  15 2 11.4870 34.9411 2.04 .
SHP611  16 2  6.3543 30.5729 2.04 .
;
run;


/************ 2. Optimal Full Matching **********/
/*
&macro variables:
   dist = Distance(Mahalanobis Distance)
   maxmatch = Maximum number of control(treated) units to be matched
                 with each treated(control) unit (KMAX, KMAXTREATED)
   fullwgt = ATT Weight
*/

/*Example of a specific Scenario*/
%let fullwgt = MATCHATTWGT;
%let dist = mah(var=(Duration OnsetAge)/cov=IDENTITY);
%let maxmatch = 5;
/* End Example Scenario */

ods output MatchInfo     =MatchInfo;
ods output StdDiff       =StdDiff;
proc psmatch data=sample region=allobs;
   class study BaseGMFC;
   psmodel study(treated="SHP611")= Duration OnsetAge;
   match distance=&dist method=full(kmax=&maxmatch
kmaxtreated=&maxmatch)
          exact=(BaseGMFC) caliper(mult=one)=.; *Minimize distance
within each exact GMFC category and no caliper is used;
   assess ps lps var=(BaseAge OnsetAge)/ VARINFO weight=&fullwgt;
   output out(obs=match)=match_full matchid=_MatchID weight=_wght;
run;


/********* 3. Interval Censored Analysis (p-value) ************/

/*
Currently there is no procedure available to use weights for interval
censored analysis.
We plan to use Strata obtained from optimal full matching
*/

ods output HomTests=ht_stats HomStats=HomSta;
proc iclifetest data= match_full plots=survival(strata=panel)
impute(seed=1234);
   strata _MatchID;
   time (lt, rt);
   test study;
```

```
run;

*** sign is obtained from Generalized Log-Rank Statistic and test
statistic calculated from one-sided normal ***;
data HT;
   set ht_stats; set homsta(where=(study ="SHP611"));
   sign = sign(stat)*-1;
   sqrt_chi = sqrt(ChiSq);  **square root of chi-square statistic;
   if sign >=0 then zstat = sign*sqrt_chi;
   os_pvalue = 1 - probnorm(zstat);  **one sided p-value from normal;
   RejectH0_1s = (os_pvalue < 0.025);
   if sign <0 then RejectH0_1s =0;
   RejectH0 = (ProbChiSq < 0.05);  **keep original stratified test for
comparison purpose;
run;

/*Print one-sided p-value*/
proc print data = HT;
title "One-sided p-value";
var os_pvalue;
run;
title;


/********* 4. Interval Censored Analysis (effect size) ************/
/* This is an approach for the estimation of the treatment effect
(using the Kaplan-Meier survival curve). The p-value obtained using
this process must not be used.  */

/*
The survival curve is not dependent on weights. The weight multiplier
should be chosen (in order to use the 'freq' statement in SAS) such
that it converts all raw weights into integers. Since the maximum
number of control/treated subjects to be matched within each stratum is
set at 5, a general weight multiplier of 60 can be used (lowest common
multiple of 2, 3, 4 and 5). The below SAS code will yield interval
survival probability estimates for each treatment group. For the
quantification of uncertainty in these survival probabilities, the
corresponding standard error output can be used, but will first need to
be converted due to the use of weight multiplier: standard error
without weight multiplier = sqrt(weight multiplier)*standard error with
weight multiplier).

*/


data match_full;
   set match_full;
   frq = round(_wght*60);
run;

proc iclifetest data=match_full plots=(survival) impute(seed=1234);
   freq frq;
```

```
time (lt, rt);
test study;
run;
```

10.5 Analysis Software

SAS Version 9.4 or higher will be used for programming and analysis of data. R may be used for specific analyses and the corresponding details would be provided in programming specifications.

10.6 Mathematical Formulation for Mean Adjusted Multidimensional Euclidean Distance

To determine the index date in GLIA-MLD, an objective selection process will be applied for each subject in GLIA-MLD who has qualified as an external control for Group A subject with multiple 'qualifying encounters':

(1) For each qualifying encounter of GLIA-MLD external control subjects, calculate the mean adjusted multidimensional Euclidean distance from qualifying encounters of all Group A subjects with the same GMFC-MLD category at the Screening visit in Study SHP611-201. Here the distance is based on the 2 continuous covariates including:

   i) age at MLD symptom onset, and
   ii) duration from onset of MLD symptoms to the qualified encounter,

The approach mentioned above can be explained using the following mathematical expressions:

h: cohort name (h = 1, Study SHP611 -201; h = 2, GLIA-MLD)

i: GMFC-MLD level M_i (i = 1, 2)

n_{hi}: Total number of subjects in cohort h, GMFC-MLD level M_i

j: j-th subject (j = 1, ..., n_{hi}) in cohort h, GMFC-MLD level M_i

k: potential start of TTE (k = 1, ..., K_{hij}), K_{hij} is cohort, patient specific. Especially K_{1ij} = 1.

a_{hijk}: age of onset of MLD symptoms at the candidate start of the k-th TTE for j-th subject in cohort h and GMFC-MLD level M_i

d_{hijk}: time from onset of MLD symptoms to qualifying medical care at the candidate start of the k-th TTE for j-th subject in cohort h and GMFC-MLD level M_i

The average adjusted multidimensional Euclidean distance Δ (a_{21j^*k}, d_{21j^*k}) between the external control subject j^* (with k-th "qualified encounter" and GMFC-MLD level M1), and the n_{11} Group A subjects with GMFC-MLD level 1 at the screening visit of Study SHP611-201 is:

$$\Delta\left(a_{21j^*k}, d_{21j^*k}\right) = \frac{1}{n_{11}} \sum_{j=1}^{n_{11}} \sqrt{\frac{\left(a_{21j^*k} - a_{11j1}\right)^2}{S_A^2} + \frac{\left(d_{21j^*k} - d_{11j1}\right)^2}{S_D^2}}$$

However, S_A^2 and S_D^2 (variances of the age at symptom onset and duration from onset of MLD symptoms to the qualified encounter, respectively) are as follows:

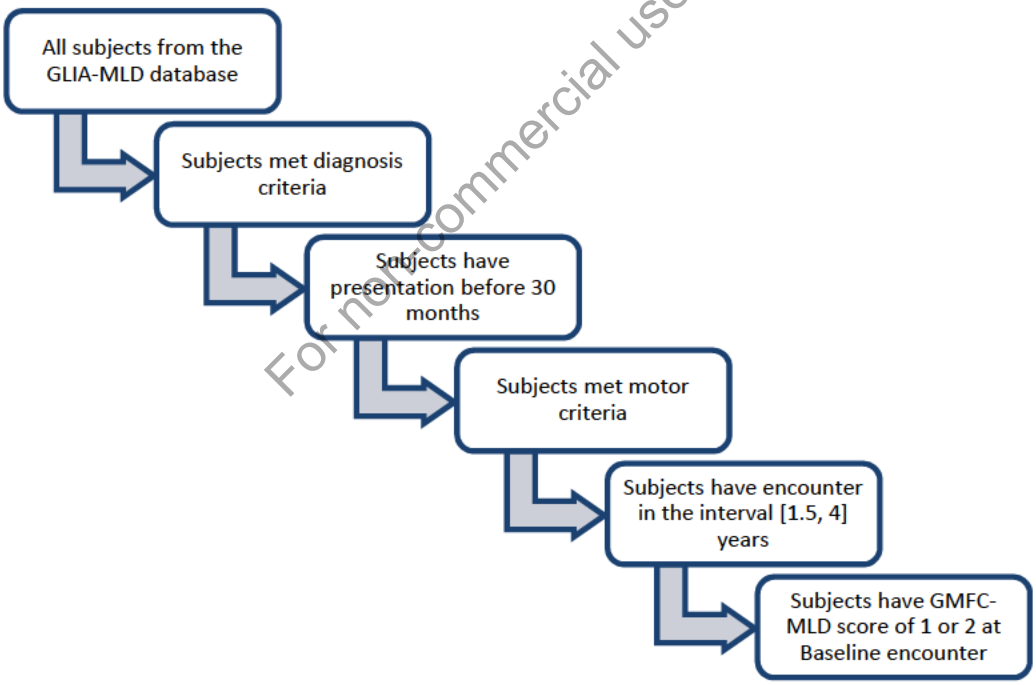
$$S_A^2 \equiv \frac{1}{\sum_{hi} n_{hi}} \sum_{hijk} \left(a_{hijk} - \frac{1}{\sum_{hi} n_{hi}} \sum_{hijk} a_{hijk} \right)^2$$

$$S_D^2 \equiv \frac{1}{\sum_{hi} n_{hi}} \sum_{hijk} \left(d_{hijk} - \frac{1}{\sum_{hi} n_{hi}} \sum_{hijk} d_{hijk} \right)^2$$

(2) The encounter associated with the lowest average adjusted multidimensional Euclidean distance is selected as the TTE starting point for the GLIA-MLD subject. In other words, for an external control subject j in GLIA-MLD at GMFC-MLD level M1, the TTE start point would be the minimum *l*th 'Eligible encounter' in the $\{\Delta(a_{21jk}, d_{21jk})\}_{k=1}^{K_{21j}}$

10.7 GLIA-MLD External Control Funneling Process

For the GLIA-MLD external control group, the number of subjects meeting each filtering criterion 1-3 (Section 3.1) will be presented through the below funneling process:



10.8 ELFC-MLD Categorical Scale

ELFC-MLD is a categorical scoring system with 5 levels:

Table 4 ELFC-MLD: Expressive Language Function Classification

ELFC-MLD Category	Description
0	Communicates in complete sentences at a quality and performance normal for age
1	Communicates in complete sentences at a reduced quality of performance for age
2	Cannot communicate in complete sentences but able to use 2-word phrases
3	Cannot communicate in 2-word phrases, but able to use single, meaningful words/ideas
4	Complete loss of expressive language

10.9 EDACS Categorical Scale

EDACS is a categorical scoring system with 5 levels:

Table 5 EDACS: Eating and Drinking Ability Classification System

EDACS Category	Description
1	Eats and drinks safely and efficiently
2	Eats and drinks safely but with some limitations to efficiency
3	Eats and drinks with some limitations to safety; there may be limitations to efficiency
4	Eats and drinks with significant limitations to safety
5	Unable to eat or drink safely, tube feeding may be considered to provide nutrition

10.10 Impact of different sensitivity analyses on the primary efficacy analysis workflow

The impact of the different sensitivity analyses on the primary efficacy analysis workflow can be summarized below:

Table 6 The impact of different sensitivity analyses on the primary efficacy analysis workflow

Sensitivity analysis number	Impact input dataset[1]?	Impact matching variable[2]?	Impact matching method[3]?	Impact matching results?	Impact analysis method[4]?
1	Yes	Yes	No	Yes	No
2a	Yes	Yes	No	Yes	No
2b	Yes	Yes	No	Yes	No
2c	Yes	Yes	No	Yes	No
2d	Yes	Yes	No	Yes	No
2e	Yes	Yes	No	Yes	No
3a	No	No	No	Yes*	No
3b	No	No	Yes	Yes	No
3c	No	No	No	Yes	No
3d	No	No	Yes	Yes	Yes
3e	No	No	Yes	Yes	Yes
4	No	N/A	N/A	N/A	Yes
5	No	N/A	N/A	N/A	Yes
6	Yes	Yes	No	Yes	No
7	Yes	No	No	No	No
8	Yes	No	No	Yes	No
9	Yes	No	No	Yes	No
10	Yes	Yes	No	Yes	No

For primary analysis:
[1] Input dataset: all patients in mFAS with variables GMFC-MLD score (GMFC-MLD category at the Screening visit and at the TTE starting point will be used for SHP611-201 and GLIA-MLD subjects, respectively), age at MLD symptom onset, duration from MLD symptom onset to TTE starting point and the TTE interval.
[2]Matching variable: GMFC-MLD score, age at MLD symptom onset, duration from MLD symptom onset to TTE starting point.
[3]Matching method: optimal full matching using Mahalanobis distance.
[4]Analysis method: interval censored generalized log-rank test.
*The strata obtained from optimal full matching will not differ, only the weights will change.