Statistical Analysis Plan

Protocol Title	An Open-Label Study of the Safety, Tolerability, and cokacokinetics of Oral NNZ-2591 in Phelan McDermid Syndrome
Protocol Number:	NEU-2591-PMS-001
Protocol Version:	2.2
SAP Version:	2.0
Date:	06 NOV 2023
Sponsor:	Neuren Pharmaceuticals Ltd. Suite 201, 697 Burke Road Camberwell, VIC 3124 Australia
Author:	[redacted]

This study will be conducted in compliance with the protocol, Good Clinical Practice and all other applicable regulatory requirements, including the archiving of essential documents.

Signatures / Approvals

[redacted] 02 December 2023 | 7:48 AM AEDT

I approve this document
02 December 2023 | 7:48 AM AEDT

[redacted]B96A098C386EDB1BAB2

Name (Print or Type) Date

Chief Science Officer

[redacted]

[redacted] 01 December 2023 | 1:14 PM PST

[redacted]DEC7E8BF81

Name (Print or Type) Date

Chief Medical Officer

[redacted] 01 December 2023 | 10:47 PM ISST

[redacted]

Name (Print or Type) Date

Vice President, Clinical Development

[redacted]

[redacted] 01 December 2023 | 12:49 PM PST

[redacted]E6968F6CED5E4CCE9105B333C034D2B2

Name (Print or Type) Date

Biostatistician II

[redacted]

Table of Contents

List of Abbreviations

ABC-2	Aberrant Behavior Checklist-2
AE	Adverse Event
AMSE	Autism Mental Status Exam
ASD	Autism Spectrum Disorder
ATC	Anatomical/Therapeutic/Chemical
BID	Twice-a-day
BMI	Body Mass Index
BPI-SF	Behavior Problems Inventory - Short Form
CGI-I	Clinical Global Impression - Improvement
CGI-S	Clinical Global Impression - Severity
CRF	Case Report Form
CSHQ	Child Sleep Habits Questionnaire
CSR	Clinical Study Report
COVID-19	Coronavirus-19
DBP	Diastolic Blood Pressure
DQ	Developmental Quotient
DSM5	Diagnostic and Statistical Manual of Mental Disorders
DSMC	Data Safety Monitoring Committee
ECG	Electrocardiogram
EDC	Electronic Data Capture
EL	Expressive Language
EOT	End of Treatment
FM	Fine Motor
FR	Fluid Reasoning
FSIQ	Full Scale Intellectual Quotient
GIHQ	Gastrointestinal Health Questionnaire
GM	Gross Motor
ICC	Intraclass Correlation Coefficient

ICH	International Conference on Harmonization
ICND	Impact of Childhood Neurological Disability
IMP	Investigational Medicinal Product
ITT	Intent-to-Treat Population
KN	Knowledge
LOCF	Last Observation Carried Forward
MB-CDI	MacArthur-Bates Communicative Development Inventory
MedDRA	Medical Dictionary of Regulatory Affairs
MSEL	Mullen Scales of Early Learning
NVIQ	Stanford-Binet nonverbal z-deviation Intelligence Quotient
QL-Disability	Quality of Life Inventory – Disability
QR	Quantitative Reasoning
RL	Receptive Language
SAE	Serious Adverse Event
SAF	Safety Population
SAP	Statistical Analysis Plan
SB5	Stanford-Binet Intelligence Scales version 5
SBP	Systolic Blood Pressure
SOC	System Organ Class
ORCA	Observer-Reported Communication Assessment
PDF	Portable Document Format
PMS	Phelan McDermid Syndrome
[redacted]	[redacted]
PP	Per-Protocol Population
PT	Preferred Term
TEAE	Treatment-Emergent Adverse Event
TLFs	Tables, Listings, and Figures
VABS-3	Vineland Adaptive Behavior Scales – 3
VIQ	Verbal Intelligence Quotient

[redacted] [redacted]

[redacted] [redacted]

WHO	World Health Organization
WM	Working Memory

1. Introduction

Neuren Pharmaceuticals Ltd. is conducting a study to investigate a new treatment for Phelan McDermid Syndrome (PMS). The study background, design and subject assessments for the study are described in the study specific protocol.

The statistical methods to be implemented during the analyses of data collected within the scope of this study will be outlined in this document. The purpose of this plan is to provide specific guidelines from which the statistical analysis will proceed. This statistical analysis plan (SAP) is to be interpreted in conjunction with the protocol. Should the SAP and protocol be inconsistent with respect to the planned analyses, the language of the SAP is governing. Any deviations from this plan will be documented in the clinical study report (CSR).

The statistical principles applied in the design and planned analyses of this study are consistent with the International Conference on Harmonisation (ICH) guidelines E9 (Statistical Principles for Clinical Trials).

1.1. Study Objectives

The primary objective of this study is to investigate the safety, tolerability, and pharmacokinetics of treatment with NNZ-2591 Oral Solution, 50 mg/mL Investigational Medicinal Product (IMP) in children and adolescents with Phelan McDermid Syndrome.

The secondary objective of this study is to investigate measures of efficacy during treatment with NNZ-2591 Oral Solution, 50 mg/mL Investigational Medicinal Product (IMP) in children and adolescents with Phelan McDermid Syndrome.

1.2. Study Design

This is a Phase 2, open-label study of the safety and tolerability of NNZ-2591 Oral Solution, 50mg/mL Investigational Medicinal Product (IMP) in male and female children and adolescents with Phelan McDermid Syndrome. Approximately 20 male and female subjects between the ages of 3 and 12 years will be enrolled and receive treatment of NNZ-2591 Oral Solution, 50mg/mL IMP for a total of 13 weeks. The target dose will be attained following 6 weeks of up-titration.

There will be a total of 17 study visits, comprising 5 in-clinic visits and 12 remote telemedicine/in-home nurse visits. The study will commence with an approximately 4-6 week Screening and Baseline period. During this Screening/Baseline period, subjects will be assessed for study eligibility and data will be collected to establish the subject's baseline characteristics and symptom severity using a variety of assessments. Two

in-clinic visits and one remote/in-home visit (Visits 1, 2, 3) comprise Screening and Baseline.

Once eligibility is confirmed, subjects will be dosed with the starting dose of 4 mg/kg and then be up-titrated to the target dose 12 mg/kg BID. Subjects will receive IMP for a total of 13 weeks. During the treatment period, there are three in-clinic visits: Week 2 (Visit 5), Week 6 (Visit 6) and Week 13 (Visit 16). There is a combined telemedicine/in-home visit by the site Investigator and an in-home visit by a visiting nurse for safety at Week 1 (Visit 4), Week 4 (Visit 7), Week 8 (Visit 11) and Week 10 (Visit 13). These combination telemedicine/in-home visits will be conducted simultaneously whenever possible, or they will be conducted on the same day with the telemedicine visit occurring first if possible. In-home visits with a visiting nurse will be done at Weeks 3 (Visit 6), 5 (Visit 8), 7 (Visit 10), 9 (Visit 12), 11 (Visit 14), and 12 (Visit 15).

All subjects will also have a combined telemedicine/in-home follow-up visit with the Investigator via telemedicine and an in-person home visit by a visiting nurse approximately 2 weeks after the end of treatment (Week 15, Visit 17).

For all visits designated as in-clinic, an in-person visit is preferred. Off-site assessments may be allowed due to extenuating circumstances due to the COVID-19 health emergency. All requests for off-site assessments must be approved in advance by the Sponsor or Medical Monitor. For visits that are designated as in-home/telemedicine visits (e.g. the combined remote visit or nurse-only in-home visit), these may be conducted in-clinic by the study staff if agreed upon by the Investigator and the caregiver.

Subjects will be divided into three groups by age for enrollment:

- 9 to 12 years old (Group 1)
- 6 to 8 years old (Group 2)
- 3 to 5 years old (Group 3)

Enrollment will commence with the oldest age group. After at least three subjects in Group 1 have received two weeks of treatment at the starting dose, the Data Safety Monitoring Committee (DSMC) will review data on safety and tolerability. If tolerability and safety for those subjects during the specified period is deemed acceptable, enrollment for Group 2 will proceed and dosing will begin at the starting dose. When at least three subjects in Group 2 have received the first two weeks of treatment at the starting dose, the DSMC will review data on safety and tolerability. If tolerability and safety for Group 2 is deemed acceptable, enrollment for Group 3 will commence and dosing will begin at the starting dose.

A summary of the schedule of events and assessments is provided in Table 1 below.

Table 1: Schedule of Events and Assessments for Neu-2591-PMS

Period	Screening/ Baseline[a] 4 to 6 weeks		Baseline	Treatment Period													Follow-up[b]
Visit Week	**-4 to -6**	**Visit 1 +2 wks**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13/EOT[b]**	**15**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Visit window (days)	**±2[a]**	**±1[a]**	**±2[a,c]**	**±1**	**±2**	**±1**	**±1**	**±1**	**±2**	**±1**	**±1**	**±1**	**±1**	**±1**	**±1**	**-3**	**+4**
Type of Visit	**Clinic[d]**	**Remote and In-Home[e,f]**	**Clinic[d]**	**Remote and In-Home[e,f]**	**Clinic[d]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-home[f]**	**Clinic[d]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-Home[f]**	**In-home[f]**	**Clinic[d]**	**Remote and In-Home[e,f]**
Informed consent (mandatory)	X																
Consent for reuse of leftover Biomarker and PK samples (optional)	X																
Inclusion/exclusion criteria	X		X														
Medical history	X	X	X														
Confirm documented PMS diagnosis and genotype	X																
PMS history and exam	X																
Confirm comorbid psychiatric disorders by DSM-5	X																
Rapid SARS-CoV-2 test	X		X													X	
Autism Mental Status Exam	X																
Recent clinical history				X	X		X		X		X		X			X	X
Physical examination[g]	X		X		X		X[h]		X		X[h]		X[h]			X	X[h]
Targeted Physical Exam[g]		N		N		N	N	N		N	N	N	N	N	N		N

Period	Screening/ Baseline[a]		Baseline	Treatment Period													Follow-up[b]
	4 to 6 weeks																
Visit Week	-4 to -6	Visit 1 +2 wks	0	1	2	3	4	5	6	7	8	9	10	11	12	13/EOT[b]	15
Visit Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Visit window (days)	±2[a]	±1[a]	±2[a,c]	±1	±2	±1	±1	±1	±2	±1	±1	±1	±1	±1	±1	-3	+4
Type of Visit	Clinic[d]	Remote and In-Home[e,f]	Clinic[d]	Remote and In-Home[e,f]	Clinic[d]	In-Home[f]	Remote and In-Home[e,f]	In-home[f]	Clinic[d]	In-Home[f]	Remote and In-Home[e,f]	In-Home[f]	Remote and In-Home[e,f]	In-Home[f]	In-home[f]	Clinic[d]	Remote and In-Home[e,f]
Neurological examination	X		X		X		X[h]		X		X[h]		X[h]			X	X[h]
[redacted]																	
Ophthalmic examination[i]	X				X				X							X	
Vital signs	X	N	X	N	X	N	N	N	X	N	N	N	N	N	N	X	N
Height/length	X		X													X	
Weight	X	N	X	N	X	N	N	N	X	N	N	N	N	N	N	X	N
Review for AEs/SAEs[j]				X, N	X	N	X, N	N	X	N	X, N	N	X, N	N	N	X	X, N
Caregiver Diary[k]	X	X	X	X, N	X	N	X	N	X	N	X	N	X	N	N	X	X, N
IMP administration[c]			X	Twice daily													
Review dosing compliance documentation in diary[l]				X, N	X	N	X, N	N	X	N	X, N	N	X, N	N	N	X	
12-lead ECG	X		X		X				X							X	N
CBC with differential	X		X		X		N		X		N		N			X	N
Coagulation[m]	X		X		X		N		X		N		N			X	N
Comprehensive metabolic panel[n]	X		X	N	X	N	N	N	X	N	N	N	N	N	N	X	N
Urinalysis[o]	X		X	N	X	N	N	N	X	N	N	N	N	N	N	X	N
Urine screen for drugs of abuse[o]	X																
TSH, Free T3, Free T4	X															X	
HbA1c	X															X	

Period	Screening/ Baseline[a] 4 to 6 weeks		Baseline	Treatment Period													Follow-up[b]
Visit Week	**-4 to -6**	**Visit 1 +2 wks**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13/EOT[b]**	**15**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Visit window (days)	**±2[a]**	**±1[a]**	**±2[a,c]**	**±1**	**±2**	**±1**	**±1**	**±1**	**±2**	**±1**	**±1**	**±1**	**±1**	**±1**	**±1**	**-3**	**+4**
Type of Visit	**Clinic[d]**	**Remote and In-Home[e,f]**	**Clinic[d]**	**Remote and In-Home[e,f]**	**Clinic[d]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-home[f]**	**Clinic[d]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-Home[f]**	**In-home[f]**	**Clinic[d]**	**Remote and In-Home[e,f]**
Serum pregnancy test[p]	X								X							X	
Blood samples for PK[q]					X				X							X	
Blood sample for biomarkers[r]	X		X													X	
Stool sample for microbiome	X		X													X	
CGI-I									X							X	X
CGI-S	X		X						X							X	X
[redacted]	[redacted]		[redacted]						[redacted]							[redacted]	
Stanford-Binet Intelligence Scales version 5 (SB5) or Mullen Scales of Early Learning[r]	X															X	
MB-CDI			X													X	
ORCA			X													X	
Caregiver Top 3 Concerns	X	X	X						X							X	X
ABC-2	X	X	X						X							X	X
BPI-SF			X													X	
Caregiver Global Impression-Change																X	
QI-Disability			X													X	

Period	Screening/ Baseline[a] 4 to 6 weeks		Baseline	Treatment Period													Follow-up[b]
Visit Week	**-4 to -6**	**Visit 1 +2 wks**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13/EOT[b]**	**15**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Visit window (days)	**±2[a]**	**±1[a]**	**±2[a,c]**	**±1**	**±2**	**±1**	**±1**	**±1**	**±2**	**±1**	**±1**	**±1**	**±1**	**±1**	**±1**	**-3**	**+4**
Type of Visit	**Clinic[d]**	**Remote and In-Home[e,f]**	**Clinic[d]**	**Remote and In-Home[e,f]**	**Clinic[d]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-home[f]**	**Clinic[d]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-Home[f]**	**Remote and In-Home[e,f]**	**In-Home[f]**	**In-home[f]**	**Clinic[d]**	**Remote and In-Home[e,f]**
ICND-Overall quality of life rating			X													X	
CSHQ	X	X	X						X							X	X
Vineland Adaptive Behavior Scales-3			X													X	
GIHQ	X	X	X						X							X	X
Study Exit Form																	X[s]

X = Conducted by site personnel or caregiver as applicable

N = Conducted by visiting nurse

Abbreviations: ABC-2=Aberrant Behavior Checklist-2; AE=adverse event; BPI-SF=Behavior Problems Inventory-Short Form; CBC=Complete Blood Count; CGI-I=Clinical Global Impression Scale-Improvement; CGI-S=Clinical Global Impression Scale-Severity; CMP=Comprehensive Metabolic Panel; CSHQ=Child Sleep Habits Questionnaire; DSM-5=Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; ECG = electrocardiogram; EOT=end of treatment; GIHQ=GI Health Questionnaire; ICND=Impact of Childhood Neurological Disability; IMP=Investigational Medicinal Product; MB-CDI=MacArthur-Bates Communicative Development Inventory; ORCA=Observer-Reported Communication Assessment; PK = pharmacokinetic; PMS=Phelan-McDermid syndrome; [redacted] QI-Disability=Quality of Life Inventory- Disability; SAE=serious adverse events; Stanford Binet Intelligence Scales ver 5; T3 = triiodothyronine; T4 = thyroxine; TSH = thyroid stimulating hormone.

[a] No assessments can be conducted until after informed consent has been given by the caregiver or legally authorized representative. Training for the caregivers on all caregiver completed assessments must be done before the caregiver/rater may complete the assessments. The Screening/Baseline period will be conducted over 4 to 6 weeks. The second visit during the Screening/Baseline period (Visit 2) should be conducted 2 weeks ± 1 day after the Screening visit (Visit 1). The subject will then remain in Screening for an additional 2 to 4 weeks. The caregiver diary must be completed daily during the entire Screening/Baseline period whether it is 4 weeks or 6 weeks. The Screening visit (Visit 1) and Baseline visit (Visit 3) may be conducted over three consecutive days. It is preferrable that ECGs, safety labs and biomarker sample collection are conducted on the same day. All assessments must be completed within the window for the visit (visit day ± 2 days). For Visit 3, all assessments must be completed before administration of the IMP.

[b] The End of Treatment visit is conducted when the subject is still on IMP treatment. For early termination, in addition to study completers, subjects who have received IMP for longer than 2 weeks and discontinue prematurely any time after Visit 5 should return to the investigational site for final safety and efficacy assessments as scheduled for the EOT visit (Visit 16) and complete the study exit form.

[c] At Visit 3, the first dose of IMP will be administered in the clinic after all Baseline assessments are completed, or, if the Investigator judges that it is too late in the day, on the following day. The subject should be monitored for at least three hours post first dose. If the first dose is administered the day following Baseline, vital signs (except height) should be collected before the dose is administered. Day 1 of dosing is defined by the date when the first dose is taken. At the Baseline visit, the site will dispense an adequate supply of IMP for dispensing by the caregiver at home. Dosing at the end of Baseline counts as Day 1 whether one or two doses are given. Dosing may not exceed 91 days including Day 1, and all efficacy assessments for the EOT visit must be completed on or before Day 91 of treatment.

[d] A clinic visit is preferred but an off-site visit may be conducted for a scheduled in-clinic visit in extenuating circumstances due to the COVID-19 health crisis. All requests for off-site visit assessments due to extenuating circumstances must be approved in advance by the Sponsor or Medical Monitor. The in-clinic visits may be conducted over three consecutive days. All assessments including caregiver-completed assessments must be done within the window for the visit.

[e] This is a combined remote and in-person nurse visit. The remote visit is conducted by the site via telemedicine or video facilities as appropriate. Any caregiver-completed assessments should be completed the same day of the visit or no earlier than the day before the visit (except the Caregiver Diary which is completed daily) and reviewed by the Investigator. A modified version of the physical and neurological exam will be done at the remote visits based on the guidance document provided to the Investigator. At Week 4, the site should confirm with the caregiver that the child has started the new dose if they were approved for dose titration. These combination remote/in-home visits will be conducted simultaneously whenever possible, or they will ideally be conducted on the same day. If that is not possible, the telemedicine visit and the nurse visit may be conducted over two consecutive days. If that is not possible, the telemedicine visit and the nurse visit may be conducted over two consecutive days. A combined remote visit may be conducted in-clinic by the study staff if agreed upon by the Investigator and the caregiver. See Section 9.1 and Section 9.2.

[f] In-home visits will be conducted by a visiting nurse. At every nurse visit, the nurse will collect urine, vital signs including weight, and blood samples for a comprehensive metabolic panel (CMP). The nurse will also conduct a targeted physical exam assessing general appearance, abdominal/CVA tenderness, and edema. A suprapubic exam (bladder palpation) to assess for urinary retention may be performed. The visiting nurse will also review the dosing diary, query oral intake, and do a review of tolerability. Blood samples for CBC with differential and coagulation will be collected at Weeks 4, 8, 10 and Week 15 (indicated by "N" in the Schedule of Assessments). At the post-treatment follow-up visit at Week 15, the visiting nurse will also collect ECG, in addition to the above assessments. At Week 7, the site will also have a phone call with the caregiver to confirm that the child has started the new dose if they were approved for dose titration. The site will confirm the visiting nurse has been informed so the dosing schedule can be reviewed with the family during the in-home visit. A nurse-only in-home visit may be conducted in-clinic by the study staff if agreed upon by the Investigator and the caregiver. See Section 9.1 and Section 9.2.

[g] The in-clinic physical exam will include an abdominal exam and an assessment of **CVA.** A suprapubic exam (bladder palpation) to assess for urinary retention may be performed. General awareness will be captured as part of the neurological exam. For in-home nursing visits, the nurse will do a Targeted Physical exam assessing general appearance (e.g., alertness), an abdominal/CVA, and edema. A suprapubic exam (bladder palpation) to assess for urinary retention may be performed.

[h] A modified version of the neurological and physical exam will be done for the remote visit based on the guidance document provided to the Investigator.

[i] [redacted]

[j] The visiting nurse will query any changes in tolerability during the in-home visit. Any changes will be noted and provided to the Investigator. The Investigator will also query changes in tolerability and note any changes in the remote assessment. The Investigator is responsible for determining and reporting any changes that qualify as Adverse Events.

[k] The caregiver fills out the diary daily from the first day of screening through two weeks after the EOT Visit. X denotes review of the Caregiver Diary including the daily summary log, medication log, non-medicinal treatments log, seizure log and new/changes in symptoms log. All visits: The Investigator (or appropriately credentialed designee) is responsible for reviewing for ALL visits. In-home visits: At the in-home nurse visit, the nurse will review the Caregiver Diary to flag and check any issues that could be reported as AEs. The nurse will notify the site about any urgent issues the day of the visit and note anything else in the source documents. The nurse does not sign the diary. The source documents and the diary will be sent to the site. The Investigator (or appropriately credentialed designee) will review the diary and source documents collected at the in-home visit.

Note: There are two rows for Caregiver Diary review procedures (to emphasize the importance of thorough and complete dosing documentation): this row (Caregiver Diary) denotes that the caregiver fills out the diary daily and a designated clinician reviews it for safety/tolerability; a second row (Review dosing compliance documentation in diary, see footnote "l") is intended to ensure that the caregiver is administering IMP/documenting dose properly in the Caregiver Diary.

[l] The daily dosing log in the Caregiver Diary will be reviewed at all post-baseline visits through EOT. If any doses are missed or mis-dosed, this should be queried with the caregiver and appropriate feedback given to the family. This feedback should be documented in the source documents. At the in-home nurse visit, the nurse will review the caregiver diary to check any issues and provide feedback to the family at the visit. The nurse will notify the site about any urgent issues the day of the visit and note anything else in the source documents. The nurse does not sign the daily dosing log/summary page. The source documents and the diary will be sent to the site. The Investigator (or appropriately credentialed delegee) will review the diary and source documents collected at the in-home visit.

[m] Coagulation includes prothrombin time, activated partial thromboplastin (aPTT), and international normalized ratio (INR).

[n] A Comprehensive Metabolic Panel will be assessed every post-enrollment visit (see Section 9.5.2). If abnormal findings are observed on renal or liver function tests, additional tests as is clinically indicated may be conducted as described in Sections 9.5.2.1 and 9.5.2.3).

[o] Both urinalysis and a urine screen for drugs of abuse will be evaluated at Screening (See Section 9.5.2).

[p] For female subjects who have reached menarche.

[q] See **Error! Reference source not found.** and **Error! Reference source not found.** for PK blood collection times.

[r] The Mullen is done for subjects who cannot obtain basal on SB5 or for whom the SB5 is developmentally inappropriate.

[s] The study exit form should also be completed if the subject terminates the study early.

1.2.1. Study Treatment

The total administration period for NNZ-2591 Oral Solution, 50 mg/mL Investigational Medicinal Product (IMP) will be 13 weeks. Following up-titration, the IMP will be administered open-label at the target dose of 12 mg/kg twice a day (BID). As outlined below in Table 2, the IMP will be administered at a starting dose of 4 mg/kg BID for 2 weeks. Following DSMC review and approval of each dose titration, subjects will be up-titrated to 8 mg/kg BID for 2 weeks, and then to the 12 mg/kg BID dose. The DSMC will review data after the subject has been dosed for 2 weeks on the 12 mg/kg BID dose to confirm safety. The subject will remain on the 12 mg/kg dose (or highest tolerated dose) for a total of 7 weeks (Table 2).

Table 2: Dose Titration Schedule

Week 1 and Week 2	Week 3	Week 4 and Week 5	Week 6	Week 7 and Week 8	Week 9	Week 10-Week 13
4 mg/kg BID (8 mg/kg daily)	4 mg/kg BID (8 mg/kg daily) [1,2]	8 mg/kg BID (16 mg/kg daily)	8 mg/kg BID (16 mg/kg daily) [1]	12 mg/kg BID (24 mg/kg daily)	12 mg/kg BID (24 mg/kg daily)	12 mg/kg BID (24 mg/kg daily)
	DSMC review data prior to up-titration [2]		DSMC review data prior to up-titration [2]		DSMC reviews data after 2 weeks on 12 mg/kg to confirm safety	

[1]Subjects will stay on their current dose until the data is reviewed by the Data Safety Monitoring Committee (DSMC)
[2]Expected time for DSMC review will be one week.

Subjects who are unable to tolerate the study treatment may be discontinued from the study.

2. Analysis Populations

Intent to Treat Population (ITT): The Intent to Treat population will consist of all subjects enrolled into the study.

Safety Population (SAF): The Safety population will consist of all subjects in the ITT population who have received at least one dose of IMP.

Pharmacokinetic Population (PK): The PK population will include all subjects enrolled in this study who receive the IMP through to the morning dose of Week 2 (Visit 5) as a minimum, and who undergo PK sample collection at at least one of the specified post-dose time point(s).

The pharmacokinetic analyses will be described in a separate analysis plan and reported separately.

3. Data Monitoring

An independent DSMC will monitor the progress of the trial and ensure that the safety of trial subjects is not compromised. The DSMC will review the clinical safety data prior to any individual dose titration and prior to initiation of subsequent age groups. The DSMC will review safety data and make a determination regarding 1) the ability for any individual subject to escalate to the next higher dose and 2) the ability for enrollment to begin in the younger age groups (Groups 2 and 3). The DSMC will consist minimally of a clinical chair, another physician experienced in clinical trials with pediatric expertise (but not participating in this study), and a pediatric cardiologist. The DSMC Charter details the DSMC processes, duties, and responsibilities as well as the study-wide stopping criteria.

Data for the DSMC is provided as raw data listings for up-titration decisions and opening of enrollment in Groups 2 and 3, as specified in the DSMC Charter. Overall safety reviews are also conducted as defined in the DSMC charter and data is provided as SAS listings. Additional summary analyses may be done as requested. There is no separate Statistical Analysis Plan for the DSMC.

4. Statistical Methods

The statistical analyses will be reported using summary tables, listings, and figures (TLFs). Numbering for TLFs will be based on the recommended numbering convention provided by the International Conference on Harmonization. Both prespecified and *a posteriori* exploratory analyses will be conducted consistent with the exploratory design of the study. Unless noted otherwise, all statistical tests will be two-sided with a nominal significance level of $\alpha = 0.05$. Continuous variables will be summarized with means, standard deviations, medians, minimums, and maximums. Categorical variables will be summarized by counts and percent of subjects in corresponding categories. Missing values are not considered for percent calculations, unless stated otherwise. Footnotes will be used to specify the percent calculation when applicable. Baseline is defined in Section 4.3 for efficacy and safety parameters.

Individual subject data obtained via the electronic data capture (EDC) system and from external vendors will be presented in listings.

The analyses described in this plan are considered *a priori*, in that they have been defined prior to database lock. Any analysis added after the database lock will be considered post hoc and exploratory. Post hoc analyses will be labeled as such on the output and identified in the CSR. All analyses and tabulations will be performed using SAS version 9.4 or higher. Tables, listings, and figures will be presented in portable document format (PDF). Upon completion, all SAS programs used to produce datasets and tables will be validated by an independent programmer. The validation process will be used to confirm that statistically valid methods have been implemented and that all data manipulations and calculations are accurate. Checks will be made to ensure accuracy, consistency with this plan, consistency within tables and consistency between tables and corresponding data listings.

4.1. Determination of Sample Size

Sample sizes were estimated for this open-label study using within-subject change as the measure of improvement. Given the limited data from interventional trials in these patient populations, and that this is the first clinical study of NNZ-2591 IMP in this population, treatment change estimates were based on Clinical Global Impression Scale of Improvement (CGI-I) data from completed clinical trials in other rare, neurodevelopmental disorders. Based on the information provided in Glaze et al. 2019 (Double-blind, randomized, placebo-controlled study of trofinetide in pediatric Rett syndrome), the standard deviation of the CGI-I is approximately 0.7 at a single timepoint. For a sample size of 20 subjects, a two-sided significance level of 0.05, and a Wilcoxon signed-rank test, the study will be able to detect a difference of 0.55 units of change in the CGI-I at a power of 0.90. For a power of 0.80, 0.85, and 0.90, the detectable difference would be 0.48, 0.51, and 0.55 or larger, respectively. The effect size for a power of 0.80, 0.85, and 0.90 is 0.68, 0.73, and 0.79, respectively. If the observed standard deviation is larger, the size of the detectable difference would also be larger. For example, if the standard deviation is 1.45, the detectable difference for a power of 0.80, 0.85, and 0.90 is 1.02, 1.09, and 1.18 or larger, respectively.

4.2. Handling of Dropouts or Missing Data

4.2.1. Missing Dates

For incomplete dates related to adverse events, concomitant medications, or medical history, the dates will be imputed as follows:

> If the incomplete date is a start/onset date:
> (1) if the month and year are present, then the first day of the month will be used for day.

(2) if only the year is present, then the first day of January will be used for month and day.

If the incomplete date is an end date:
(1) if the month and year are present, then the last day of the month will be used for day.
(2) if only the year is present, then the last day of December will be used for month and day. If the reported year is the same as the informed consent year, then the informed consent date will be used.

If the date of last dose is missing, then it will be imputed to the subject participation end date.

Dates that are completely missing will not be imputed.

Dates related to efficacy measures will not be imputed.

4.2.2. Baseline Values

Baseline values that are missing will not be imputed for safety or efficacy measures. For assessments collected at more than one visit during the Screening/Baseline period,the baseline measurement for an efficacy variable will be based on the average all non-missing values from visits 1-3.

4.2.3. Post-baseline Values

For subjects who discontinue the study, data will be analyzed up to the last observation for safety and efficacy measures. Missing data for safety assessments will not be imputed. For efficacy measures, if an assessment has been completed but has missing items, missing values shall be imputed as follows:

- If 3 or fewer items are missing, the last observation for that particular data element (e.g., item response) will be carried forward.
- The scoring algorithms will be applied including the observed values and last observation carried forward (LOCF) for the missing items
- These rules will only apply to individual items. Total scores, subdomain and subscale scores will not be carried forward. If the assessment is completely missing, items will not be imputed.

4.3. Definitions

- The baseline measurement for an efficacy variable will be based on the average values of visits 1-3 for assessments collected at more than one visit. For assessments collected once during the Screening/Baseline period, the most recent non-missing assessment will be used. For safety assessments, value at Visit 3 (Baseline) or the most recent non-missing assessment will be used.

- Age will be derived as the integer value of informed consent/assent date minus birth date plus one, then divided by 365.25.

- The age at time of assessment in calculated as (date of assessment – date of birth + 1)/365.25.

- Body Mass Index (BMI) will be derived as weight in kilograms divided by height in meters squared.

- The change from baseline measurement for a variable at a particular visit will be derived as the measurement taken at the specified visit minus the baseline measurement. CGI-I will not have a change from baseline measurement; the within-subject treatment change is assessed by the value at the End of Treatment Visit.

- Treatment duration in days will be derived as follows: Date of Last Dose – Date of First Dose + 1.

- Study duration will be defined as the total number of days in the study (date of last visit – date of informed consent + 1).

- History of developmental regression will be determined from the medical history question as reported on the PMS diagnosis page of the CRF: “Has the child undergone a regression of skills in the past?” If coded as “yes” the subject is considered as having had a past developmental regression.

- Autism Spectrum Disorder (ASD) diagnosis will be derived from the PMS diagnosis page of the CRF. If ASD is checked to the under “confirmation of psychiatric co-morbidities”, the subject is considered as having the disorder.

- Bipolar Disorder will be derived from the PMS diagnosis page on the Case Report Form (CRF). If “Bipolar disorders” is checked to the under “confirmation of psychiatric co-morbidities”, the subject is considered as having the disorder.

- Attention Deficit Disorder will be derived from the PMS diagnosis page on the CRF. If "Attention Deficit Disorder" is checked to the under "confirmation of psychiatric co-morbidities", the subject is considered as having the disorder.

- Obsessive-compulsive Disorder will be derived from the PMS diagnosis page on the CRF. If "Obsessive-compulsive disorder" is checked to the under "confirmation of psychiatric co-morbidities", the subject is considered as having the disorder.

- Dose modification is defined as dose increased, dose reduced, or dose interrupted responses for the action taken with study treatment on the adverse event (AE) CRF.

- Improvement in the top 3 concerns will be derived as a decrease in severity (change from baseline <0).

- Baseline seizure frequency will be determined based on the number of seizures recorded in the diary during the 4–6-week screening period. The baseline monthly seizure frequency will be derived by dividing the total number of seizures during the screening/baseline period by the screening/baseline duration in months ([first dose date – screening date] divided by 30.4375). The baseline weekly seizure frequency will be derived by dividing the total number of seizures during the screening/baseline period by the screening/baseline duration in weeks ([first dose date – screening date] divided by 7). During the treatment period (Visits 4 – 16), the seizure frequency will be derived by summing the number of seizures recorded in the diary over treatment. The monthly seizure frequency will be derived by dividing the total number of seizures during the treatment period by the treatment duration in months (treatment duration divided by 30.4375). The weekly seizure frequency will be derived by dividing the total number of seizures during the treatment period by the treatment duration in weeks (treatment duration ([first dose date – screening date] divided by 7).

- The Mullen Scales of Early Learning developmental quotient (DQ) is calculated as the (mean of the age equivalent in months of the corresponding scales/chronological age in months) x 100. For the overall DQ, the mean of the fine motor, receptive language, expressive language, visual reception is included. For the non-verbal DQ, the mean of the fine motor and visual reception scales are included. For the verbal DQ, the receptive language and expressive language scales are included.

- For the Mullens Scales of Early Learning and the Stanford-Binet 5, Chronological age in months will be calculated as (date of assessment – date of birth + 1)/30.4375.

- The number of concomitant medications will be summed across subjects for all unique medications that are taken while on treatment. Medications taken with a frequency of PRN or "as needed" will be excluded from the count.

5. Subject Information

5.1. Subject Disposition

Information regarding subject disposition will be summarized for all subjects by treatment group. Summaries will include the following: number of subjects enrolled, number of subjects in the ITT, Safety, and PK population, number of subjects completing the study, and number of subjects who discontinue the study early. For those who discontinue early, the primary reason for discontinuation will be summarized. A data listing by subject will also be generated.

5.2. Protocol Deviations

Protocol deviations will be collected throughout the duration of the study. Protocol deviations will be assigned a sponsor-defined category type. Each deviation will be defined as Important or non-important. In addition, a by-subject listing of all protocol deviations (important and non-important) will be produced.

5.3. Demographic and Baseline Characteristics

Demographic variables will include the following: age, sex, ethnicity, race, height at baseline, weight at baseline, and BMI at baseline. Age and BMI will be derived as specified in Section 4.3.

Baseline characteristics will include: PMS genotypes (as recorded on the CRF), history of regression, Autism Mental Status Exam, seizure clinical types at screening (as recorded on the Seizures and Spells Log and Seizure Types at Screening CRFs), presence of co-morbid psychiatric disorders (e.g. Autism Spectrum Disorder (ASD), bipolar disorders, attention deficit disorder, obsessive-compulsive disorder), and Stanford-Binet Intelligence Scales version 5 (SB5) non-verbal (NVIQ) z-deviation score or the Mullen Scales of Early Learning non-verbal developmental quotient for subjects who cannot achieve basal level on the SB5 or for whom the SB-5 is developmentally inappropriate.

Demographics and baseline characteristics and will be summarized for all analysis populations (ITT Population, Safety Population, PK population).

Subjects will be evaluated for PMS at the Screening Visit. The diagnosis and genotypes will be documented. PMS genotypes will be summarized and listed by subject.

History of developmental regression will be determined from the medical history question as reported on the PMS diagnosis page in the CRF: "Has the child undergone a regression of skills in the past?" If coded as "yes" the subject is considered as having had a past developmental regression.

The Autism Mental Status Exam (AMSE) will be performed at the Screening Visit. The AMSE is an 8-item observational assessment used to document communicative and behavioral functioning in subjects with ASD. Each question is scored as 0, 1, or 2. The range of the AMSE score is 0 – 14. Higher scores indicate greater severity of autistic symptoms.

The scoring algorithm is below:

1. Question 1-4, Eye Contact, Interest in Others, Pointing Skills, Language: each response is assigned a score of 0, 1, or 2 in the order they appear on the CRF.
2. Question 5 Pragmatics of Language: Not Impaired or Not Applicable=0, Reported=1, Observed=2
3. Question 6 Repetitive Behaviors/Stereotypy: None=0, Insists on routines=1, any other response=2
4. Question 7-8 Unusual or Encompassing Preoccupations, Unusual Sensitivities: None=0, Reported=1, Observed=2

The AMSE score is the sum of the individual question scores above. The AMSE score and the overall impairment will be summarized.

Seizure type ("Clinical Type" in Seizures and Spells Log form on the CRF recorded at Baseline) and characteristics of seizures ("Description" in Seizure Types at Screening form on the CRF: change in awareness, loss of urine or bowel control, loss of ability to communicate, automatic repeated movements, rhythmic jerking muscle, falls if standing or sitting, head drops, warning before seizure occurred, common triggers, muscle stiffness, muscle twitch, other) are recorded at the Screening Visit for subjects. These descriptive characteristics will be summarized using frequencies and percentages.

The presence of other co-morbid psychiatric disorders will be recorded on the PMS diagnosis page of the CRF. If the disorder is checked on the CRF, it will be considered as present in the subject.

5.4. Medical History

Medical history will be coded using MedDRA version 25.1 and listed by System Organ Class (SOC), Preferred Term (PT), and verbatim term. A complete medical history will

be recorded at all Screening and Baseline visits (Visits 1, 2, 3). Medical history will be summarized using frequencies and percentages according to the SOC and PT.

6. Safety Analysis

The primary objective of the study is to investigate the safety, tolerability, and pharmacokinetics of treatment with NNZ-2591 in children and adolescents with PMS.

All safety analyses will be based on the Safety Population.

Safety will be assessed based on incidence of treatment emergent adverse events (TEAEs), serious adverse events (SAEs), changes in vital signs, changes in laboratory measures, neurological, ophthalmological, and physical exam results, renal ultrasound findings, and changes in ECG's. In addition to the outputs (tables, listings, shift tables) described below for each safety variable, other visual presentations of the data may be generated (e.g., volcano plots) as appropriate.

Tolerability will be assessed based on daily dosing logs, required dosing adjustments or discontinuations, dosing compliance, and AEs related to reported subject experience.

Pharmacokinetics will be summarized separately outside of this SAP.

6.1. Dosing Summary

Treatment duration and study duration will be derived as stated in Section 4.3 and summarized using descriptive statistics.

The maximum dose (4 mg/kg BID, 8 mg/kg BID, or 12 mg/kg BID) of each subject will be summarized according to the Dosing Compliance CRF as a frequency distribution. The number of missed doses, as recorded in the Daily Dosing Log CRF, will be summarized.

6.2. Adverse Events

Adverse events summaries will only consider treatment emergent adverse events (TEAEs). TEAEs are defined as those adverse events that occurred after dosing and those existing adverse events that worsened during the study based on the start date and time. If it cannot be determined whether the adverse event is treatment emergent due to an incomplete (partial) onset date, the adverse event will be considered treatment emergent. AEs will be recorded from the time of first dose of IMP until Follow-up (Visit 17).

Verbatim terms entered into the clinical database via the EDC system will be mapped to preferred terms and system organ classes using version 25.1 of Medical Dictionary for Regulatory Activities (MedDRA).

Summaries that are displayed by system organ class and preferred terms will be ordered by descending order of incidence of system organ class and preferred term within each system organ class. Adverse event summaries of the following types will be presented for the Safety Population:

- Overall summary of the TEAEs which contain an overview of each item below.
- Subject count and incidence rate of TEAEs and total number of unique TEAEs by MedDRA system organ class and preferred term.
- Incidence rate of TEAEs by MedDRA preferred term and highest severity. Severity is defined by three categories: “Mild”, “Moderate”, and “Severe.” At each level of subject summarization, a subject is classified according to the highest severity if the subject reported one or more events. Adverse events with missing severity will be considered severe for this summary.
- Subject count and incidence rate of TEAEs by MedDRA preferred term and closest relationship to study drug (Related/Not Related). At each level of subject summarization, a subject is classified according to the closest relationship if the subject reported one or more events. Adverse events with missing relationship will be considered related for this summary.
- Subject count and incidence rate of Serious TEAEs by MedDRA system organ class and preferred term.
- Subject count and incidence rate of TEAEs leading to withdrawal of study medication by MedDRA system organ class and preferred term.
- Subject count and incidence rate of TEAEs leading to dose modification (defined as dose increased, dose reduced, dose interrupted on the CRF) by MedDRA system organ class and preferred term.
- Subject count and incidence of TEAEs leading to death as an outcome by MedDRA system organ class and preferred term.
- Subject count and incidence rate of TEAEs of special interest (defined as any renal event as indicated on the AE CRF, and all other AESI (ophthalmological, cardiac, and liver function) will be determined by a clinical review of MedDRA coded terms prior to database lock) by MedDRA system organ class and preferred term.

6.3. Clinical Laboratory Evaluations

Blood and stool samples will be collected at Visit 1 (Screening), Visit 3 (Baseline), and Visit 13 (End of Treatment) or upon early termination for analysis of biomarkers and gut

microbiome. Biomarkers and gut microbiome analysis will be described in a separate analysis plan.

Blood and urine samples will be collected at all except Visit 2 for screening and safety. Laboratory parameters (chemistry, hematology, urinalysis) will be presented in data listings. Abnormal values will be flagged as high or low relative to normal ranges, where applicable, by the laboratory vendor.

Change from baseline in numeric laboratory parameter values will be summarized using descriptive statistics at all visits.

For those laboratory tests in which normal ranges are available, shifts from baseline will be presented. Categories for the shift tables will be low, normal and high.

Laboratory parameters will be classified according to the categories given by the laboratory vendor. The laboratory categories from the vendor listed below will be assigned to the following categories for data summaries.

Chemistry	Hematology	Urinalysis
Amphetamines Conf Bun/Creatinine Ratio EGFR (Bedside Schwartz) Hemoglobin A1C Pregnancy, qualitative Thyroxine, Free Triiodothyronine, Free TSH Ultrasensitive Urine Drug Screen w/ Con	PT/INR/APTT	Microalbumin, UR Random Urine Chemistry

6.4. Physical and Neurological Examination [redacted]

A physical examination will be conducted at all in-clinic visits. The procedure will include the following organ systems: Head, ears, eyes, nose, and throat; Skin; Cardiovascular; Respiratory; Gastrointestinal; Genitourinary (as appropriate or indicated); Musculoskeletal; and Allergies. The exam will include an assessment for sign of bladder or kidney pathology including an exam of the abdomen and both flanks.

Assessment of edema will also be done. Exam may include suprapubic exam (bladder palpation) if decreased urine output is noted.

A neurological examination will also be conducted at all in-clinic visits. A standard neurological exam will be conducted to include an evaluation of mental status, cranial nerves, motor exam, sensory exam, cerebellar (coordination, gait), and reflexes.

A modified version of the physical and neurological exams will be done by a physician via telemedicine at the remote visits. As part of the in-home visit, the nurse will conduct a targeted physical exam including assessment of general appearance, an abdominal and a costovertebral angle tenderness exam, assessment for signs of cystitis, and an assessment of edema. Exam may include suprapubic exam for bladder palpation if decreased urine output is noted.

[redacted]

All examination results will be reported as normal or abnormal. Clinical significance will be determined for abnormal results.

The results of all examinations will be presented in data listings.

6.5. Ophthalmology Assessment

[redacted] The complete exam conducted at Visit 1 and Visit 16 will include 2 attempts at [redacted] a dilated fundus examination.

The results will be presented in data listings.

6.6. Vital Signs

Weight, systolic blood pressure (SBP; in mmHg), diastolic blood pressure (DBP; in mmHg), heart rate (bpm), respiratory rate (rpm), and body temperature (°C) will be collected at the Screening visits (Visit 1 and 2) and all post-baseline study visits. Vital signs will be taken after the subject has been seated for 5 minutes. The initial blood

pressure reading will be summarized in the table. Height will be measured at the Screening visit (Visit 1), Baseline (Visit 3), and End of Treatment (Week 13, Visit 16).

Summary statistics for each measurement and the change from baseline will be presented at each post-baseline visit. All vital sign measurements will be presented in data listings. A separate table and listing will be provided for blood pressure.

6.7. Electrocardiogram (ECG)

The Electrocardiogram (ECG) assessment will measure PR Interval (msec), QRS Duration (msec), QT Interval (msec), QTcF Interval (msec), QTcB Interval (msec), Heart Rate (bpm), and RR Interval (msec). The assessment will be done at the Screening visit (Visit 1), Baseline (Visit 3), Visit 5 (Week 2), Visit 9 (Week 6), End of Treatment (Week 13, Visit 16) and Follow-Up (Visit 17). The interpretation of the results will be assessed as Normal or Abnormal, and any Abnormal results will be assessed for clinical significance. The subject should be supine for at least 5 minutes before the first ECG is taken. If this cannot be done, this will be documented in the CRF. The ECG will be obtained as a continuous ECG of at least 50 ms. All readable ECGs of at least 10 ms will be interpreted by the central reader and reported.

Summary statistics for each ECG parameter and the change from baseline will be presented at each post-baseline visit. Additionally, summary statistics for the shifts in interpretation from baseline will be presented at each post-baseline visit. All ECG measurements will be presented in data listings.

6.8. Prior and Concomitant Medications

Prior and concomitant medication verbatim terms captured via the EDC system will be mapped to Anatomical/Therapeutic/Chemical (ATC) class and Preferred Names using the September 1, 2022 version of the World Health Organization (WHO) Drug Dictionary Enhanced available.

Prior and concomitant medications will be summarized separately by WHO ATC class and preferred name. Medications with an end date prior to first dose will be considered prior medications. Medications that are taken during while on study medication will be considered concomitant medications. These summaries will present the number and percent of subjects using each medication. Subjects may have more than one medication per ATC class and preferred name. At each level of subject summarization, a subject is counted once if one or more medications at that level is reported for the subject. Each summary will be ordered by descending order of incidence of ATC class and preferred term.

6.9. Total, Monthly, and Weekly Seizure Episodes

Evaluation of total, monthly, and weekly seizure frequency will be derived as defined in Section 4.3. Seizure frequency will be assessed from a seizure diary that is part of the Caregiver Diary which is completed daily by the caregiver.

Summary statistics on the observed and the change from baseline in total and monthly seizures will be displayed at baseline and end of treatment. Within subject changes will be assessed using a Wilcoxon Signed Rank Test.

7. Efficacy Analysis

The secondary objective of the study is to investigate measures of efficacy during treatment with NNZ-2591 in children and adolescents with PMS.

All efficacy analyses will be conducted on the ITT population.

7.1. Efficacy Variables

There are no primary efficacy variables.

Secondary variables for efficacy to be considered include:

- PMS-specific Clinical Global Impression Scale–Overall Improvement (CGI-I) Score
- PMS-specific Clinical Global Impression Scale–Domain Improvement Scores
- [redacted]
- PMS Domain Specific Rating Scale Improvement Scores
- Caregiver Impression of Change Score
- Clinical Global Impression Scale–Severity (CGI-S)–Overall Severity Score
- Clinical Global Impression Scale–Severity (CGI-S)–Domain Scores
- MacArthur-Bates Communicative Development Inventory (MB-CDI) Scores
- Observer-Reported Communication Ability (ORCA) Score
- Caregiver Top 3 Concerns Likert Scale Scores
- Aberrant Behavior Checklist-2 (ABC-2) Scores
- Behavior Problems Inventory-Short Form (BPI-SF) Scores
- Child Sleep Habits Questionnaire (CSHQ) Scores
- Gastrointestinal Health Questionnaire (GIHQ) Scores
- Vineland Adaptive Behavior Scales-3 Growth Scale Scores
- Quality of Life Inventory-Disability (QI-Disability) Scores
- Impact of Childhood Neurological Disability (ICND) Overall quality of life rating
- Stanford-Binet raw, z-deviation, age-equivalent, and Change Sensitive scores or Mullen Scales of Early Learning raw, age-equivalent and Developmental Quotient scores

Biomarker variables and gut microbiome analyses will be described in a separate analysis plan.

7.2. Adjustments for Covariates

A list of potential covariates to be used in descriptive summaries or analyses are provided below.

- Age (9-12, 6-8, 3-5 yrs)
- Sex (Male, Female)
- SB5 NVIQ or MSEL NV Developmental Quotient
- History of developmental regression (yes, no)
- ASD diagnosis (yes, no)

7.3. Examination of Subgroups

No subgroup analyses are planned.

7.4. Multicenter Studies

All sites will be pooled for this analysis.

7.5. Multiplicity Comparison/Multiplicity

No hypothesis testing is planned; therefore, there will be no adjustments for multiplicity and all reported significance levels will be nominal.

8. Efficacy Analysis Methods

8.1.1. Primary Efficacy Analysis

There is no primary efficacy endpoint.

8.1.2. Secondary Efficacy Analysis

The efficacy endpoints will be analyzed using a Wilcoxon signed rank test on the paired subject data from baseline to end of treatment (Visit 16) when applicable. Descriptive statistics will be used to summarize the data. For total scores, waterfall plots, histograms, and violin plots will also be created at the End of Treatment (EOT) visit with plots for all subjects and separate plots color-coded to indicate age group (9-12 years, 6-8 years, 3-5 years), NVIQ/DQ score (≥ 50, < 50), sex (Male, Female) and number of concomitant medications (> 6, ≤ 6). The number of concomitant medications will be determined as defined in Section 4.3. For scales with domain scores, forest plots will be generated.

[redacted]

Results for each efficacy variable will also be listed by subject.

8.1.2.1. Clinical Global Impression Scale—Overall Improvement (CGI-I) Score

Clinical Global Impression of Improvement assessments are completed at Visit 9 (Week 6), End of Treatment (Week 13, Visit 16), and Follow-up (Visit 17). The clinician rates how much the subject's illness has improved or worsened relative to the baseline visit (Visit 3). The Overall CGI-I score is rated on a 7-point scale, where 1='very much improved'; 2='much improved'; 3='minimally improved'; 4='no change'; 5='minimally worse'; 6='much worse'; 7='very much worse'. Scores of 1, 2, and 3 indicate improvement. Within-subject change is indicated by the score at each relevant follow-up visit (Week 6, Visit 9 and Week 13, EOT, Visit 16). If an overall score is missing, that subject will be excluded from analysis.

Summary statistics for the CGI-I Overall score will be displayed at all post-baseline visits. Within-subject changes will be analyzed using a Wilcoxon signed rank test. Waterfall plots, histograms, and violin plots will be generated as described in Section 8.1.2.

8.1.2.2. Clinical Global Impression Scale—Domain Improvement Scores

Clinical Global Impression of Improvement assessments are completed at Visit 9 (Week 6), End of Treatment (Week 13, Visit 16), and Follow-up (Visit 17). The clinician rates

how much the subject's illness has improved or worsened relative to the baseline visit (Visit 3) in each domain. The seven domain improvement scores are: expressive communication, receptive communication, gross motor function, fine motor function, social interaction, self-care, and cognition/learning. The individual domain CGI-I scores are rated on a 7-point scale, where 1='very much improved'; 2='much improved'; 3='minimally improved'; 4='no change'; 5='minimally worse'; 6='much worse'; 7='very much worse'. Scores of 1, 2, and 3 indicate improvement.

Within subject change is indicated by the score at each relevant follow-up visit (Week 6, Visit 9 and Week 13, EOT, Visit 16). If a domain score is missing, that subject will be excluded from analysis for that domain. Domain scores will be presented in forest plots, as described in Section 8.1.2.

Domain scores as reported on the CRF will be summarized at each post-baseline visit, in the same manner as the CGI-I Overall Score.

8.1.2.3. [redacted]

[redacted]

[redacted]

8.1.2.4. Caregiver Impression of Change Score

The Caregiver Impression of Change assessment will be completed at the end of treatment visit (Week 13, Visit 16).

The caregiver is asked to rate the change in his or her child's overall function and well-being since the Baseline visit, and the change in specific symptoms. This is rated on a 7-point Likert scale: 1-Very much improved, 2-Much improved, 3-Improved, 4-Unchanged, 5-Worse, 6-Much worse, 7-Very much worse. Scores of 1, 2 or 3 indicate improvement in overall function and symptoms.

For the rating of overall function, the caregiver also identifies the one symptom area or feature that has most influenced his or her rating of the child's overall function and provides a description of symptom changes and impact of the changes. The assessment of change should be made since the Baseline visit (Visit 3). The caregiver should assess the subject's current state at the visit taking into consideration observations at the visit and symptoms within the last week.

The symptom domains rated are: communication, social interaction, behavior, motor abilities, seizures, cognitive abilities and ability to learn, support needed to perform daily

tasks, gastrointestinal (GI) symptoms, and sensory sensitivity problems. Within subject change is indicated by the score at EOT. If a score is missing for a domain, that subject will be excluded from the corresponding domain summaries. For the total score, the LOCF method will be used for the missing score and the total score will be calculated.

Summary statistics on the Caregiver Impressions of Improvement overall and domain scores will be displayed at End of Treatment. The overall score will be presented in waterfall plots, histograms, and violin plots, and the domain scores will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.5. Clinical Global Impression Scale—Severity (CGI-S)—Overall Severity Score

Clinical Global Impression of Severity assessments will be made at the Screening visit (Visit 1), Baseline (Visit 3), Visit 9 (Week 6), End of Treatment (Week 13, Visit 16), and Follow-up (Week 15, Visit 17). The overall severity score is scored on a 7-point scale, where 1=’not at all impaired’; 2=’borderline, slightly impaired’; 3=’mildly impaired’; 4=’moderately impaired’; 5=’markedly impaired’; 6=’severely impaired’; 7=’among the most severely impaired’. Lower scores reflect less impairment. Within subject change is the difference between the score at EOT and the score at Baseline (Visit 3). If a score is missing, the subject will be excluded from the corresponding summaries.

The overall score and change from baseline at each post-baseline visit will be summarized using descriptive statistics. Within-subject changes will be analyzed using a Wilcoxon signed rank test. Test-retest reliability will be assessed for all scores collected during the Screening period (Visits 1, 2, and 3, pre-dosing) using Intraclass Correlation Coefficients (ICC). The overall score will be presented in waterfall plots, histograms, and violin plots, as described in Section 8.1.2.

8.1.2.6. Clinical Global Impression Scale—Severity (CGI-S)—Domain Scores

Clinical Global Impression of Severity assessments will be made at the Screening visit (Visit 1), Baseline (Visit 3), Visit 9 (Week 6), End of Treatment (Week 13, Visit 16), and Follow-up (Week 15, Visit 17). The seven domain severity scores are: expressive communication, receptive communication, gross motor function, fine motor function, social interaction, cognition/learning and self-care. Each domain score is scored on a 7-point scale, where 1=’not at all impaired’; 2=’borderline, slightly impaired’; 3=’mildly impaired’; 4=’moderately impaired’; 5=’markedly impaired’; 6=’severely impaired’; 7=’among the most severely impaired’. Lower scores reflect less impairment. Within subject change is the difference between the total score at EOT and the score at Baseline (Visit 3). If a score is missing, the subject will be excluded from the corresponding summaries.

The CGI-S domain scores will be analyzed in the same manner as the CGI-S overall score.

8.1.2.7. MacArthur-Bates Communicative Development Inventory (MB-CDI) Scores

The MacArthur-Bates Communicative Development Inventory (MB-CDI) will be completed at Baseline (Visit 3) and End of Treatment (Week 13, Visit 16). This caregiver-completed instrument assesses language development in vocabulary comprehension and production, gestures, and grammar. Designed to assess early language development in infants and toddlers, it can also be used with older, developmentally delayed children to assess developmental progress in language development. Children are assessed on one of three modules based on chronological or developmental age: Words and Gestures (8 to 18 months or developmental equivalent), Words and Sentences (16 to 30 months or developmental equivalent) and the CDI-III, an extension of the MB-CDI used with children 30 to 37 months (or developmental equivalent). The appropriate module will be used for this study in line with the cognitive and language development of the individual subject.

In the Words and Gestures module, the number of phrases understood, the number of words understood, number of words produced, number of early gestures, number of later gestures, and number of gestures will be summarized. The number of "yes" answers in "First signs of understanding" and "Starting to Talk" will be summarized. In the Words and Sentences module, the number of words produced, number of word forms, number of word endings, and the number of times a more complex sentence is selected will be summarized. In the EDC, for "words produced" in the WS module, a duplicate empty field was available in the EDC. For the analysis, the data entered in field variable name MBCDI2_CDINUM is used for analysis. For the CDI-III module, the total number of words recognized in the Vocabulary Checklist, the number of responses where the more complex sentence was chosen in Sentences, and the number of "Yes" responses in Using Language will be summarized. All other MB-CDI results will be presented in listings.

Change from baseline for the scores from the module completed for the child will be analyzed using a Wilcoxon signed rank test. To summarize data across the modules, the data on vocabulary will be presented for all subjects in a summary table. This will include the relevant scores from which ever module was completed: WG (the number of words understood); WS (number of words produced), CDI-III (total number of words for vocabulary). The vocabulary score will be presented as waterfall plots, histograms, and violin plots as described in Section 8.1.2. Exemplars of new words may be listed for individual subjects who show increases in vocabulary.

8.1.2.8. Observer-Reported Communication Ability (ORCA) Score

The Observer-Reported Communication Ability (ORCA) measure will be completed at Baseline (Visit 3) and End of Treatment (Week 13, Visit 16). This caregiver-reported measure was developed to assess communicative abilities in verbal and low-verbal/non-verbal populations. The caregiver evaluates the individual's observed communication ability over the past 30 days. The ORCA measure will be scored using the R code provided by the developers of the measure. The macro will produce an ORCA t-score as long as at least one of the 20 concepts or items has a score. A concept score will be given only if all the items within the concept are answered. Higher scores reflect greater communication ability. The ORCA t-score range is 26.82 to 83.24.

A modified t-score will be also calculated. The SAS macro used to calculate the total score will be edited to give credit for both "always" and "sometimes" responses. This edited macro will be used to calculate the modified t-score.

The t-score and modified t-score will be summarized using descriptive statistics. Change from baseline will be analyzed using a Wilcoxon signed rank test. The t-score and modified t-score will be presented in waterfall plots, histograms, and violin plots, as described in Section 8.1.2.

8.1.2.9. Caregiver Top 3 Concerns Rating

The Caregiver Top 3 Concerns Rating assessment will be completed at all Screening and Baseline visits (Visits 1, 2, 3), as well as Week 6 (Visit 9), End of Treatment (Week 13, Visit 16), and Follow-up (Visit 17). The three concerns identified at Screening/Baseline are carried forward and rated at the follow-up visits. The severity of each concern is scored using a 10-point Likert scale of severity. Lower scores indicate lesser severity. The overall score will be calculated as a sum of the scores across all domains. The overall score will have a range of 0 to 30.

If the caregiver does not list the same 3 concerns on the Top 3 Concerns Follow-Up CRF that were recorded on the Top 3 Concerns CRF (completed at Visit 1), then their responses will be reviewed and may not be included in analysis. A review of all concerns will be done prior to database lock, and visits with incorrect concerns will be flagged as unusable for analysis.

The frequency of any symptom being chosen by concern number, along with all concerns combined will be summarized. Summary statistics for the change from baseline in the overall and the concern scores and by symptom domain will be displayed at all post-baseline visits. Test-retest reliability will be assessed for all scores collected during the Screening period (Visits 1, 2, and 3, pre-dosing) using Intraclass Correlation Coefficients

(ICC). The overall and concern scores will be presented in waterfall plots, histograms, and violin plots, and the domain scores at baseline will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.10. Aberrant Behavior Checklist-2 (ABC-2)

The Aberrant Behavior Checklist-2 (ABC-2) will be completed at all Screening and Baseline visits (Visits 1, 2, 3), as well as Week 6 (Visit 9), End of Treatment (Week 13, Visit 16), and Follow-up (Visit 17). Each item is scored on a 4-point scale, where 0= Not at all a problem; 1= The behavior is a problem but slight in degree; 2= The problem is moderately serious; 3= The problem is severe in degree. The items are grouped into 5 subscales: Irritability (Subscale I, score ranges from 0 – 45), Social Withdrawal (Subscale II, score ranges from 0 – 48), Stereotypic Behavior (Subscale III, score ranges from 0 – 21), Hyperactivity/Noncompliance (Subscale IV, score ranges from 0 – 48), and Inappropriate Speech (Subscale V, score ranges from 0 – 12). The total score has a range from 0 – 174. Lower scores indicate fewer behavior issues. The subscale scores were calculated via a scoring algorithm in the EDC. These calculated scores will be used for the analysis. Missing items will be imputed as defined in section 4.2.3. Subscale or total scores may be calculated from answered and imputed items. If a total or subscale score is missing, then that subject will be excluded from the corresponding summaries.

Within subject change is the difference between the score at each measured timepoint and the score at Baseline (Visit 3). Summary statistics for the change from baseline in total and each subscale ABC-2 score will be displayed at all post-baseline visits. Change from baseline in the total and subscale scores will be analyzed using the Wilcoxon signed rank test. Test-retest reliability will be assessed for all scores collected during the Screening period (visits 1, 2, and 3, pre-dosing) using Intracass Correlation Coefficients (ICC). The total score will be presented in waterfall plots, histograms, and violin plots, and the subscale scores will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.11. Behavior Problems Inventory—Short Form (BPI-SF)

The Behavior Problems Inventory (BPI-SF) will be completed at Baseline (Visit 3) and at the End of Treatment (Week 13, Visit 16). The BPI-SF is a caregiver-completed rating scale for assessing self-injury, aggressive and stereotyped behavior. Each item is measured for frequency and for severity for the self-injurious behavior and aggressive/destructive behavior subscales. The frequency is scored on a 5-point scale, where 0=never; 1=monthly, 2=weekly, 3=daily, 4=hourly. The severity is scored on a 3-point scale, where 1=mild, 2=moderate, 3=severe. The items are grouped in three subscales: Self-Injurious Behavior (score ranges from 0 – 32 for frequency and 8 – 24 for severity), Aggressive/Destructive Behavior (score ranges from 0 – 40 for frequency and

10 – 30 for severity), and Stereotyped Behavior (score ranges from 0 – 48 for frequency). The score for each domain is the sum of the items within that domain. The total frequency score is the sum of the frequency score for the three domains The total severity score is the sum of the severity scores for aggressive/destructive behavior and self-injurious behavior. Missing items will be imputed as defined in section 4.2.3. A domain score will only be calculated if all items in that domain are answered or can be imputed. Lower scores indicate lesser frequency and severity in behavior problems. Within subject change is the difference between the score at EOT and the score at Baseline (Visit 3).

Summary statistics for the total and domain scores and their changes from baseline in the frequency and severity of the domain-specific BPI-SF scores will be displayed at End of Treatment. Within subject changes will be analyzed using a Wilcoxon signed rank test. The frequency and severity scores will be presented in waterfall plots, histograms, and violin plots, and the domain scores will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.12. Child Sleep Habits Questionnaire (CSHQ)

The Child Sleep Habits Questionnaire (CSHQ) will be completed at all Screening and Baseline visits (Visits 1, 2, 3), as well as Week 6 (Visit 9), End of Treatment (Week 13, Visit 16), and Follow-up (Visit 17). Each item is scored on a 3-point scale, where 1=rarely; 2=sometimes; 3=usually. Items 32 and 33 are scored as 1 = not sleepy, 2 = very sleepy and 3 = Falls asleep A higher score reflects more disturbed sleep behavior. The items are grouped into 4 domains and 8 subscale scores are derived from the 33 items in these 4 domains. The subscale scores are a sum of the items within that subscale, except for a few items which are reversed for scoring. Missing items will be imputed as defined in section 4.2.3. The subscale scores will only be calculated if all items in subscale are answered or can be imputed. The subscales and their corresponding items are listed below.

1. Bedtime Resistance: Items 1, 3, 4, 5, 6, 8, where items 1 and 3 are reversed. Subscale score ranges from 6 – 18.
2. Sleep Onset Delay Behavior: Item 2 where item 2 is reversed. Subscale score ranges from 1 – 3.
3. Sleep Duration: Items 9, 10, 11, where items 10 and 11 are reversed. Subscale score ranges from 3 – 9.
4. Sleep Anxiety: Items 5, 7, 8, 21. Subscale score ranges from 4 – 12.
5. Night Wakings: Items 16, 24, 25. Subcale score ranges from 3-9.
6. Parasomnias: Items 12, 13, 14, 15, 17, 22, 23. Subscale score ranges from 7-21.
7. Sleep Disordered Breathing : Items 18, 19, 20. Subscale score ranges from 3-9.

8. Daytime Sleepiness: Items 26, 27, 28, 29, 30, 31, 32, 33, where item 26 is reversed. Domain score ranges from 8 – 24.

The total score will be calculated as a sum of all 33 items, with items 1 – 3, 10 – 11, and 26 reversed. The total score will range from 33 – 99 and will only be calculated if all items are answered. Within-subject change is the difference between the score at each timepoint and the score at Baseline (Visit 3).

Summary statistics for the derived scores and their changes from baseline in the total sleep disturbance score and domain-specific CSHQ scores will be displayed at all post-baseline visits. Test-retest reliability will be assessed for all scores collected during the Screening period (Visit 1, 2, and 3, pre-dosing) using Intraclass Correlation Coefficients (ICC). Changes from baseline will be analyzed using a Wilcoxon signed rank test. The total score will be presented in waterfall plots, histograms, and violin plots, and the subscale scores will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.13. Gastrointestinal Health Questionnaire (GIHQ) Scores

The Gastrointestinal Health Questionnaire (GIHQ) will be completed at all Screening and Baseline visits (Visits 1, 2, 3), as well as Week 6 (Visit 9), End of Treatment (Week 13, Visit 16), and Follow-up (Visit 17). Each item is scored on its frequency, relevancy, and importance to the caregiver. The Frequency is measured on a 5-point scale where 0=never; 1=almost never; 2=sometimes; 3=often; 4=almost always. The surgery domain, however, records frequency as 0=No and 1=Yes. The relevancy is scored on a scale from 1 to 4, where 1 is not relevant and 4 is very relevant. The importance is scored on a scale from 1 to 4, where 1 is not important and 4 is very important. Lower scores indicate fewer gastrointestinal problems. The total score is the sum of all items. The items are grouped into 9 domains. The domain score is the sum of all items within the domain, as shown on the CRF. Missing items will be imputed as defined in section 4.2.3. The domain scores will only be calculated if all items within the domain are answered or can be imputed. The domains are as follows:

1. General Health/Pain: Items 1 – 5, frequency score ranges from 0 – 20, relevancy score ranges from 5 – 20, importance score ranges from 5 – 20.
2. Eating, Chewing, and Swallowing: Items 1 – 9, frequency score ranges from 0 – 36, relevancy score ranges from 9 – 36, importance score ranges from 9 – 36.
3. Reflux: Items 1 – 3, frequency score ranges from 0 – 12, relevancy score ranges from 3 – 12, importance score ranges from 3 – 12.
4. Gas and Bloating: Items 1 – 5, frequency score ranges from 0 – 20, relevancy score ranges from 5 – 20, importance score ranges from 5 – 20.

5. Diarrhea and Constipation: Items 1 – 6, frequency score ranges from 0 – 24, relevancy score ranges from 6 – 24, importance score ranges from 6 – 24.
6. Personality and Mood: Items 1 – 5, frequency score ranges from 0 – 20, Relevancy score ranges from 5 – 20, importance score ranges from 5 – 20.
7. Medications: Items 1 – 9, frequency score ranges from 0 – 36, relevancy score ranges from 9 – 36, importance score ranges from 9 – 36.
8. Surgery: Items 1 – 5, frequency score ranges from 0 – 5, relevancy score ranges from 5 – 20, importance score ranges from 5 – 20.
9. Parenting: Items 1 – 8, frequency score ranges from 0 – 24, relevancy score ranges from 8 – 24, importance score ranges from 8 – 24.

Within subject change is the difference between the score at each timepoint and the score at Baseline (Visit 3). Summary statistics for the change from baseline in the total and each GIHQ domain score will be displayed at all post-baseline visits. Test-retest reliability will be assessed for all scores collected during the Screening period (Visits 1, 2, and 3, pre-dosing) using Intraclass Correlation Coefficients (ICC). Changes from baseline will be analyzed using a Wilcoxon signed rank test. The total score will be presented in waterfall plots, histograms, and violin plots, and the domain scores will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.14. Vineland Adaptive Behavior Scales-3 (VABS-3)

The Vineland Adaptive Behavior Scales-3 will be completed at Baseline (Visit 3) and End of Treatment (Week 13, Visit 16). Scores are calculated by the site rater. An Adaptive Behavior Composite (ABC) standard score will be recorded, along with a standard score for each of the 4 domains: Communication, Daily Living Skills, Socialization, Motor Skills. For the motor skills domain, scores can only be calculated for participants ages 3-9. These range from 20 to 140.

In each of the 11 subdomains, the raw scale (range 0 to 116), v-scale (range 1 to 24), age equivalent (range 0:0 to 22:0+), and growth scale (range 10 to 197) scores are recorded. Lower scores reflect less adaptability. If a score is missing, that subject will be excluded from corresponding summaries. Within subject change is the difference between the score at EOT and the score at Baseline (Visit 3).

Summary statistics for the change from baseline in the Vineland Adaptive Behavior Scales-3 composite and domain standard scores, and subdomain raw, v-scale, and growth scale scores will be displayed at end of treatment. Changes from baseline will be analyzed using a Wilcoxon signed rank test. The composite standard score will be presented in waterfall plots, histograms, and violin plots, and the domain standard scores will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.15. Quality of Life Inventory—Disability (QL-Disability)

The Quality of Life Inventory—Disability (QL-Disability) Score will be completed at Baseline (Visit 3) and at the End of Treatment (Week 13, Visit 16). Each item is scored on a 5-point scale, where 1=never; 2=rarely; 3=sometimes; 4=often; 5=very often. The items are grouped into five domains: health and well-being (items 1-4), feelings and emotions (items 5-15), family and friends (items 16-22), activities and the outdoors (items 23- 27) and Daily Life (items 28-32). The domain scores are calculated by first transforming the 5-point scale, so that 1=0, 2=25, 3=50, 4=75, and 5=100. Items 9-15 are reversed in scoring, so that 1=100, 2=75, 3=50, 4=25, and 5=0. The domain score is the average of the transformed score of the items within that domain. Missing items will be imputed as defined in section 4.2.3. A domain score will be calculated as long as at least one of the items within the domain is answered. The overall score is the average of all domain scores and will be calculated as long as at least one of the domain scores can be calculated. Higher scores indicate a better quality of life. Within subject change is the difference between the score at EOT and the score at Baseline (Visit 3).

Summary statistics for the observed and change from baseline in the QL-Disability overall and domain scores will be displayed at end of treatment. Changes from baseline will be analyzed using a Wilcoxon signed rank test. The overall score will be presented in waterfall plots, histograms, and violin plots, and the domain scores will be presented in a forest plot, as described in Section 8.1.2.

8.1.2.16. Impact of Childhood Neurological Disability (ICND) - Overall quality of life rating

The Impact of Childhood Neurological Disability (ICND) Scale will be administered at Baseline (Visit 3) and at End of Treatment (Week 13, Visit 16). The ICND is scored on a scale from 1 to 6, where 1=poor and 6=excellent. Lower scores indicate greater negative impact on quality of life. Within subject change is the difference between the score at EOT and the score at Baseline (Visit 3). If a score is missing, that subject will be excluded from corresponding summaries.

Summary statistics for the observed and change from baseline in the ICND score will be displayed at end of treatment. Change from baseline will be analyzed using a Wilcoxon signed rank test. The overall score will be presented in waterfall plots, histograms, and violin plots, as described in Section 8.1.2.

8.1.2.17. Stanford-Binet Intelligence Scales (SB5)

The Stanford-Binet Intelligence Scales (SB5) will be performed at Screening and Visit 16 (end of treatment). Participants who cannot be assessed on the SB-5 due to their

developmental level will be assessed using the Mullen Scales of Early Learning instead. General cognitive ability is reported as a full scale intellectual quotient score (FSIQ). Cognitive ability in five categories is reported in the index scores: fluid reasoning (FR), knowledge (KN), quantitative reasoning (QR), visual-spatial processing (VS), and working memory (WM). Non-verbal (NVIQ) and verbal intelligence quotients (VIQ) are also recorded. Scores range from 40 to 160, where higher scores indicate a higher IQ. For the analysis of index scores and IQ, the z-deviation method for intellectual disability will be used (Sansone et al. 2014). Standard scores, change-sensitive scores, age equivalents and the raw scores were calculated and entered by the site rater. Z-deviation scores were derived using the algorithm in R code from the algorithm developer (Translational Psychophysiology and Assessment Laboratory (T-PAL), [redacted], MIND Institute, University of California, Davis, Sacramento California). For age equivalent values < 2, the result is entered into EDC as 1.9, for analysis. For age equivalent values > X, where X is a numeric value, the result is entered into EDC as X.1, for analysis. For example, ">33" is entered in EDC as 33.1.

Summary statistics on the observed and change from baseline in the z-deviation scores, change-sensitive scores and age equivalents for FSIQ, NVIQ, VIQ, and the domain scores: FR, KN, QR, VS, WM. Each category will be displayed at end of treatment. The FSIQ, NVIQ, and VIQ z-deviation scores will be presented in waterfall plots, histograms, and violin plots, and the domain z-deviation scores will be presented in forest plots, as described in Section 8.1.2.

8.1.2.18. Mullen Scales of Early Learning

The Mullen Scales of Early Learning (MSEL) will be performed at Screening and Visit 16 (end of treatment) on those subjects who cannot be assessed using the SB5. It assesses developmental and cognitive functioning in five categories: gross motor (GM), visual reception (VR), fine motor (FM), expressive language (EL), and receptive language (RL). All scores except the developmental quotients are calculated by the site rater. A raw score, age equivalent score and developmental quotient will be recorded for each category for all participants. As part of the analysis, an overall developmental quotient, a verbal developmental quotient, and a non-verbal developmental quotient will also be calculated as follows:

Overall Developmental Quotient:

(Mean of the age-equivalent scores in months of the fine motor, receptive language, expressive language, visual reception scales/chronological age in months) x 100

Verbal Developmental Quotient:

(Mean of the age-equivalent scores in months of the expressive and receptive language scales/chronological age in months) x 100

Non-verbal Developmental Quotient:

(Mean of the age-equivalent scores in months of the visual reception and fine motor scales/chronological age in months) x 100

Chronological age in months will be calculated as (date of assessment – date of birth + 1)/30.4375.

For each of the five domains, a t-score will be recorded for participants ≤ 68 months old. An Early Learning Composite score will be recorded a long with its band of error for participants ≤ 68 months old.

For age equivalent values < 2, the result is entered into EDC as 1.9, for analysis. For age equivalent values > X, where X is a numeric value, the result is entered into EDC as X.1, for analysis. For example, ">33" is entered in EDC as 33.1.

The domain raw scores have an average of 50 and a standard deviation of 10. The composite standard score has a mean of 100 and a standard deviation of 15. The developmental quotient scores are a ratio of developmental age to chronogical age and will have a range of 1 to 100. Higher scores indicate a higher IQ or better developmental progress. If any scores are missing, the subject will be excluded from corresponding summaries.

Summary statistics on the observed and change from baseline in the composite standard score, developmental quotient scores, age equivalent scores, and each category t-score will be displayed at end of treatment. The composite standard score and overall, verbal, and non-verbal developmental quotient scores will be presented in waterfall plots, histograms, and violin plots. The domain developmental quotient scores will be presented in forest plots, as described in Section 8.1.2.

8.1.3. Exploratory Efficacy Analysis

8.1.3.1. Biomarkers and gut microbiome

Blood and stool samples will be collected at Screening (Visit 1), Baseline (Visit 3, before dosing), and at End of Treatment (Visit 8), or upon early termination. Biomarkers and gut microbiome will be analyzed separately outside of this analysis plan.

8.1.3.2. Correlation of Efficacy Endpoints

[redacted]

[redacted]

9. Tables, Listings, and Figures

Data listings will be sorted by subject number and visit (if appropriate). All data recorded on the CRFs will be included in the listings as well as some derived information. The shells for the data listings will be primarily to show where specific data can be found. The variable spacing and overall presentation or layout of the listings may need to be altered once actual data is available.

Tables will generally be presented with the parameter in the first column followed by the summary column. Descriptive statistics such as mean, median, minimum, and maximum will be presented to one decimal place and standard deviation will be presented to two decimal places. The laboratory tables may vary for each test depending on the precision of the test. P-values will be presented to four decimal places. If a p-value is less than 0.0001 then it will be presented as <0.0001.

A list of the tables, listings, and figures will be maintained outside of this document and may be amended as needed.

10. References

Sansone SM, Schneider A, Bickel E, Berry-Kravis E, Prescott C, Hessl D. Improving IQ measurement in intellectual disabilities using true deviation from population norms. *J Neurodev Disord*. 2014;6(1):16.