Official Title: A Phase 3b Open-Label Study to Evaluate the Effect of Kuvan® on Neurocognitive Function, Maintenance of Blood Phenylalanine Concentrations, Safety, and Population Pharmacokinetics in Young Children with Phenylketonuria

NCT Number: NCT00838435

Applicant/MAH: BioMarin Pharmaceutical Inc.

Version Date: 26 April 2019

16.1.9 Documentation of Statistical Methods

105 Digital Drive

Novato, CA 94949

Statistical Analysis Plan

PKU-015

(Part 2: Neurocognitive Study)

Version 3.0, April 26, 2019

Version 2.0, April 26, 2019

Original Version, June 20, 2018

1 APPROVALS (SIGNATURE AND DATE)

Title: A Phase 3b Open-Label Study to Evaluate the Effect of Kuvan® on Neurocognitive Function, Maintenance of Blood Phenylalanine Concentrations, Safety, and Population Pharmacokinetics in Young Children with Phenylketonuria

Protocol: PKU-015 Amendment 2 (version **30Nov10)**

Date: April 26, 2019

DocuSigned by: [redacted] 30-Apr-2019 | 6:41 AM PDT
Signer Name: [redacted]
Biostatistician
Signing Reason: I am the author of this document
Signing Time: 4/30/2019 6:41:17 AM PDT
[redacted]

DocuSigned by: [redacted] 26-Apr-2019 | 10:02 AM PDT
Signer Name: [redacted] MD
Signing Reason: I approve this document
Signing Time: 4/26/2019 10:02:03 AM PDT
[redacted]

DocuSigned by: [redacted] 26-Apr-2019 |
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 4/26/2019 11:31:13 AM PDT
[redacted]

[redacted] 30-Apr-2019 | 2:12 PM BST
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 4/30/2019 [redacted]
Global Medical Affairs, St[redacted]nt & Data Analytics
[redacted]

DocuSigned by: [redacted] 26-Apr-2019 |
☐ NA
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 4/26/2019 9:03:21 AM PDT
[redacted]

DocuSigned by: [redacted] 30-Apr-2019 | 6:39 AM PDT
☐ NA
Signer Name: [redacted]
Pharmacovigilanc[redacted]
Signing Reason: I approve this document
Signing Time: 4/30/2019 6:38:59 AM PDT
[redacted]

☒ NA
Medical Writer

DocuSigned by: [redacted] 29-Apr-2019 | 8:43 AM PDT
☐ NA
Signer Name: [redacted]
Regulatory Affairs
Signing Reason: I approve this document
Signing Time: 4/29/2019 8:42:53 AM PDT
[redacted]

TABLE OF CONTENTS

2 LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
AEoSI	Adverse Event of Special Interest
ATC	Anatomical Therapeutic Chemical
Bayley-III	Bayley Scale of Infant and Toddler Development®- Third Edition
BH4	Sapropterin dihydrochloride
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
CRF	Case Report Form
CSR	Clinical Study Report
ECG	Electrocardiogram
FSIQ	Full Scale IQ
HLGT	High level group term
IDC	Index of Dietary Control
IP	Investigational Product
IQ	Intelligence Quotient
MedDRA	Medical Dictionary for Regulatory Activities
Phe	Phenylalanine
PKU	Phenylketonuria
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
TEAE	Treatment-Emergent Adverse Event

Abbreviation	Definition
WISC-IV	Wechsler Intelligence Scale for Children®-Fourth Edition
WHO	World Health Organization
WHODrug	World Health Organization Drug Dictionary
WPPSI-III	Wechsler Preschool and Primary Scale of Intelligence™-Third Edition

3 INTRODUCTION

BioMarin PKU-015 is a 2-part multicenter, open-label study designed to evaluate the safety of Kuvan and its effect on neurocognitive function, blood Phe concentration, and growth in children with PKU who are 0-6 years old at study entry. The purpose of this document is to provide specific details of the statistical analysis methods and data presentations collected in Protocol PKU-015 protocol amendment 2 (dated November 30, 2010) for the Neurocognitive Study (Part 2). This excludes the analyses for substudy 1 and substudy 2, which were outlined in a separate statistical analysis plan document.

If discrepancies exist between the text of the statistical analysis as planned in the protocol and the final SAP, a protocol amendment will not be issued and the SAP will prevail.

3.1 Objectives of Study

The primary objective of the study is:

- To evaluate the long-term efficacy of Kuvan on preserving neurocognitive function in children with PKU when treatment is initiated at 0-6 years at Screening

The secondary objectives of the study are:

- To evaluate the long-term safety of Kuvan in the study population
- To evaluate the effect of Kuvan on growth parameters in the study population

3.2 Study Design

This 2-part, multicenter, open-label study is designed to evaluate the safety of Kuvan and its effect on neurocognitive function, blood Phe concentration, and growth in children with PKU when treatment is initiated at 0-6 years of age. The study will be considered complete when approximately 45 subjects have completed 7 years of treatment.

One of the goals of the study will be to decrease blood Phe levels to reach a goal of ≤ 240 µmol/L. Subjects will meet with a dietician at each study visit to review dietary Phe intake. Dietary Phe may be added to the diet, at the discretion of the Investigator, if a blood Phe level falls below 120 µmol/L or decreased if blood Phe level is above 240 µmol/L. If a subject's blood Phe falls below 120 µmol/L, at the discretion of the investigator, a gradual increase of approximately 5-20 mg/kg of Phe supplement may be added to the diet. If a subject's blood Phe level goes above 240 µmol/L, dietary Phe may be decreased by

approximately 5 to 20 mg/kg or defined by the standard used at each treatment center, at the discretion of the investigator.

Part 1: Evaluation of Kuvan Responsiveness

Beginning at Week 0 and after all baseline assessments are completed, subjects enrolled into the study will receive 20 mg/kg Kuvan daily for 4 weeks. Subjects will return to the clinic weekly for safety assessments and blood Phe concentration measurement. Subjects should continue with their current amount of dietary Phe intake. However, if a subject's blood Phe concentration drops below 120 µmol/L, at the discretion of the Investigator, a gradual increase by approximately 5 to 20 mg/kg of daily dietary Phe supplement may be added to the diet. It is important to avoid unstable swings in blood Phe concentrations. No dose adjustments can be made during Part 1 without consultation with the Medical Monitor.

Part 2: Neurocognitive Study

Subjects who are responsive to Kuvan will continue to receive once a day doses of 20 mg/kg Kuvan orally. A Kuvan-responsive subject is defined as a subject with ≥30% average reduction in blood Phe concentration calculated by the mean of the weekly percent change from baseline in blood Phe concentration at Weeks 1, 2, 3 and 4. The baseline blood Phe concentration is calculated as the mean of the results from Screening visit and at Week 0 (prior to dosing with Kuvan). At the discretion of the Investigator, the subject's daily Phe intake may be modified if blood Phe concentration falls below 120 µmol/L during the study. Kuvan dose may also be reduced after Week 5 after consultation with the Medical Monitor, in subjects who do not tolerate the 20 mg/kg/day dose.

Neurocognitive testing will be performed within 6 weeks of determination of Kuvan responsiveness in all Kuvan-responsive subjects. The types of neurocognitive tests used (Bayley Scales of Infant and Toddler Development®-Third Edition [Bayley-III], Wechsler Preschool and Primary Scale of Intelligence™-Third Edition [WPPSI-III], and/or Wechsler Intelligence Scale for Children®-Fourth Edition [WISC-IV]) will be those appropriate for the subject's current age at assessment. Table 3.2.1 shows the neurocognitive test to be used according to subject age.

Table 3.2.1: Neurocognitive Assessments

Age Group	Infancy	Preschool		School-Age
Assessment	Bayley-III	WPPSI-III (First Form)	WPPSI-III (Second Form)	WISC-IV
Age	0 months through < 30 months	30 months through < 4 years	≥ 4 years through < 7 years	≥7 years and older
Frequency	Every 6 months ± 2 weeks	Once a year ± 2 weeks	Once a year ± 2 weeks	Every 2 years ± 2 weeks

Growth, blood Phe concentration, and safety will be monitored. These and the neurocognitive assessments will be performed every 6 to 24 months, depending on the type of assessment, until the end of the study.

3.3 Study Population

Individuals eligible to participate in this study must meet all of the following criteria:

- Parent(s) or guardian(s) willing and able to provide written, signed informed consent after the nature of the study has been explained, and prior to any research-related procedures.
- Parent(s) or guardian(s) willing and able to comply with all study procedures.
- Female subjects of childbearing potential (as determined by the investigator) and sexually mature male subjects willing to use a medically accepted method of contraception throughout the study. Female subjects of childbearing potential willing to undergo periodic pregnancy tests during the course of the study
- Established diagnosis of PKU with hyperphenylalaninemia (HPA) documented in the medical record by at least 2 blood Phe concentrations ≥ 360 µmol/L (6 mg/dL) taken at least 3 days apart
- Documented blood Phe control (defined by the standard used at each treatment center) prior to study enrollment, if applicable (eg, the subject is old enough for these data to be collected); blood Phe concentrations for subjects < 6 months old at Screening must be considered controlled and stable by the Investigator

- Willing to adhere to a prescribed Phe-restricted diet in order to maintain blood Phe concentrations within the recommended ranges established at the subject's study site
- Age 0 to 6 years, inclusive, at Screening

Individuals eligible to participate in Part 2: Neurocognitive Study must meet **all** of the following criteria:

- Completion of Week 4 visit in Part 1
- Responsive to Kuvan during Part 1, defined as a $\geq$ 30% average reduction in blood Phe concentration calculated as the mean of the weekly percent change from baseline in blood Phe concentration at Weeks 1, 2, 3, and 4
- Bayley-III or IQ test score $\geq$ 80 when tested within 6 weeks of determination of Kuvan responsiveness in Kuvan-responsive subjects

Individuals who meet any of the following exclusion criteria will not be eligible to participate in the study:

- Has known hypersensitivity to Kuvan or its excipients
- Use of Kuvan or any investigational product or investigational medical device within 30 days prior to screening, or requirement for any investigational agent prior to completion of all scheduled study assessments
- Concurrent disease or condition that would interfere with study participation or safety (eg, seizure disorder, oral steroid-dependent asthma or other condition requiring oral or parenteral corticosteroid administration, or insulin-dependent diabetes)
- Use of phosphodiesterase type 5 inhibitor, often shortened to PDE5 inhibitor (eg.sildenafil citrate, vardenafil, tadalafil, avanafil, lodenafil, mirodenafil, udenafil)
- Any condition that, in the view of the PI, places the subject at high risk of poor treatment compliance or of not completing the study
- Established diagnosis of primary tetrahydrobiopterin (BH4) deficiency
- History of organ transplantation
- Perceived to be unreliable or unavailable for study participation or to have parents or legal guardians who are perceived to be unreliable or unavailable

- Use of methotrexate or other medications that inhibit folate metabolism
- Serious neuropsychiatric illness (eg, major depression) not currently under medical control

3.4 Sample Size Determination

The sample size calculation assumes each child will have 2 post-treatment test results from the WPPSI-III and /or WISC-IV tests at least 2 years apart. The goal is to exclude a mean 2.5-point loss per year in IQ score. Given a 2-side Type I error rate of 0.05, if the mean IQ score at baseline is 100, the standard deviation is no larger than 16, and the correlation between the 2 test results is 0.8, a sample size of 45 subjects will yield at least 90% power.

To account for the possibility of dropouts and to have reasonable precision for estimating change in IQ from the time a child reaches at least age 30 months until the end of the study, approximately 60-80 Kuvan-responsive subjects will be enrolled in the study.

3.5 Blinding and Randomization Methods

This is an open-label study. No blinding and randomized treatment allocations were used in this study.

3.6 Interim Analysis

There is no formal interim analysis planned. However, an analysis of data for a subset of subjects who responded to sapropterin therapy and had completed at least 2 years in the study as of June 2012, including subjects who terminated early from the study prior to reaching their 2-year study visit, was performed. Neurocognitive function, blood and dietary Phe control, growth assessments, and adverse events were examined.

4 GENERAL ANALYSIS CONSIDERATION

Continuous endpoints will be summarized using descriptive statistics such as the number of observations, mean, standard deviation, median, minimum, maximum, 25th percentile, and 75th percentile. Categorical endpoints will be summarized by the frequency and percent.

The last available measurement or procedure performed prior to initial study drug administration will be considered the baseline measurement or procedure for analysis purpose, unless otherwise specified.

All tests will be conducted at a two-sided alpha level of 0.05, and all confidence intervals will be given at a two-sided 95% level, unless stated otherwise.

No multiplicity adjustment will be made. Data from all sites will be pooled; no by-site analysis will be performed.

4.1 Analysis Populations

The Full Analysis set (FAS) will consist of all subjects who enroll into the study.

The Enrolled population will consist of all subjects who enter Part 2 of the study and met the inclusion/exclusion criteria per CRF. This population involves subjects who respond to Kuvan and had a Bayley-III or IQ test score ≥ 80 within 6 weeks of determination of Kuvan responsiveness in Part 1. Kuvan responders consist of all subjects with at least 30% average reduction in blood Phe concentration calculated as the mean of the weekly percent change from baseline at Week 1, 2, 3 and 4 in Part 1.

The Efficacy population will consist of all subjects from the Enrolled population who have at least two WPSSI/WISC assessments.

The Safety population will consist of all subjects who have received any study drug in the FAS population.

4.2 Treatment Group Presentation

All subjects enrolled into the study will be administered 20 mg/kg Kuvan daily.

Efficacy and safety analysis results will be presented by age group at enrollment, and all ages combined except for the Neurocognitive summary. The following age groups will be used for all subjects enrolled in Part 2:

- 0 – 1 year old (0 to < 1)

- 1 - 2 years old (≥ 1 to < 2)
- 2 - 4 years old (≥ 2 to < 4)
- 4 – 6 years old (≥ 4 to < 7)

In addition, safety analysis may also be presented during Part 1 by the following group:

- Part 2 Enrolled
- Part 2 Not Enrolled
- Overall

4.3 Pooling of Data from Sites with Small Enrollment

Since the enrollment for each site is expected to be small, the analyses will not be adjusted by site or pooled sites.

4.4 Study Day Derivation

Study Day 1 is used as the reference date for the derivation of Study Day in data analyses and displays, and is defined as the first date study drug is administered in Part 1. Study Day is computed by subtracting the Study Day 1 date from the visit date and plus 1 if the visit date occurs on or after the first study drug administration date. Otherwise, Study Day will be the visit date minus the Study Day 1 date. There is no Study Day 0. For specific efficacy analyses, Study Day 1 may also be defined as the day of the first neurocognitive assessment.

4.5 Visit Window for Analysis

The baseline blood Phe concentration is defined as the mean of the screening and Week 0 (prior to dosing with Kuvan) result. Baseline measurement for other assessments is defined as the latest non-missing measurement prior to the first study drug administration in Part 1. If there are two or more measurements within an analysis visit window, the measurement that is closest to the scheduled time point will be used for analyses. If the two closest measurements to the scheduled time point are equidistant from that scheduled analysis time point, the average (or more conservative) of the two assessments will be used for analyses.

Table 4.5.1 presented below defines visit windows for the scheduled analysis time points by analysis parameter.

Table 4.5.1: Visit Windows

Assessment	Scheduled Time Point	Visit Window
Bayley-III	Baseline	First measurement date after determination of Kuvan responsiveness (Day 30 to Day 120)
	Month 6	Day 121 to Day 273
	Month 12	Day 274 to Day 456
	Month 18	Day 457 to Day 638
	Month 24	Day 639 to Day 821
WPPSI-III, WISC-IV (based on Month 2 occurrence)	Baseline	First measurement date after determination of Kuvan responsiveness at Month 2 visit (Day 30 to Day 120)
	Month 12	Day 121 to Day 548
	Month 24	Day 549 to Day 913
	Month 12*i, where i=2,3,4,5,6,7,…	Day 365*i – 181, Day 365*i + 183
WPPSI-III, WISC-IV (based on first assessment)	Baseline*	First measurement date after determination of Kuvan responsiveness
	Month 12	Day 2 to Day 548
	Month 24	Day 549 to Day 913
	Month 12*i, where i=2,3,4,5,6,7,…	Day 365*i – 181, Day 365*i + 183
Height, head circumference, physical exam	Baseline	Last measurement date prior to the initial study drug administration
	Week 4	Day 2 to 76
	Month 3	Day 77 to Day 137
	Month 6	Day 138 to Day 273
	Month 12	Day 274 to Day 456

Assessment	Scheduled Time Point	Visit Window
	Month 18	Day 457 to Day 638
	Month 24	Day 639 to Day 821
	Month 6*i, where i=2,3,4,5,6,7,…	Day round(182*i + i/2) – 91, Day round(182*i + i/2) + 91 – mod(i,2)*
Tryptophan, tyrosine	Baseline	Last measurement date prior to the initial study drug administration
	Week 4	Day 2 to 76
	Month 3	Day 77 to Day 137
	Month 6	Day 138 to Day 273
	Month 12	Day 274 to Day 548
	Month 24	Day 549 to Day 913
	Month 12*i, where i=2,3,4,5,6,7,…	Day 365*i – 181, Day 365*i + 183
Blood Phe level, Dietary Phe, Weight, Vital Signs	Baseline	Measurement dates prior or on the same day of the initial Kuvan administration
	Week 1	Day 2 to Day 11
	Week 2	Day 12 to Day 18
	Week3	Day 19 to Day 25
	Week4	Day 26 to Day 42
	Month2	First measurement date after determination of Kuvan responsiveness (Day 43 to Day 76)
	Month 3	Day 77 to Day 106
	Month 4	Day 107 to Day 137
	Month 5	Day 138 to Day167
	Month 6	Day 168 to Day 273

Assessment	Scheduled Time Point	Visit Window
	Month 12	Day 274 to Day 456
	Month 18	Day 457 to Day 638
	Month 6*i, where i=2,3,4,5,6,7,…	Day round(182*i + i/2) – 91, Day round(182*i + i/2) + 91 – mod(i,2)*
Additional Week 0 Vital signs	15 min postdose	1 min to 22 min elapsed time
	30 min postdose	23 min to 37 min elapsed time
	45 min postdose	38 min to 52 min elapsed time
	60 min postdose	53 min to 75 min elapsed time
	90 min postdose	76 min to 105 min elapsed time
	120 min postdose	106 min to 150 min elapsed time
	180 min postdose	151 min to 360 min elapsed time
Clinical laboratory tests	Baseline	Last measurement date prior to the initial study drug administration
	Week 1	Day 2 to Day 15
	Week 3	Day 16 to Day 25
	Week 4	Day 26 to Day 42
	Month 2	First measurement date after determination of Kuvan responsiveness (Day 43 to Day 76)
	Month 3	Day 77 to Day 106
	Month 4	Day 107 to Day 137
	Month 5	Day 138 to Day167
	Month 6	Day 168 to Day 273
	Month 12	Day 274 to Day 456
	Month 18	Day 457 to Day 638
	Month 6*i, where i=2,3,4,5,6,7,…	Day round(182*i + i/2) – 91, Day round(182*i + i/2) + 91 – mod(i,2)*
*Baseline refers to first assessment for this visit windowing		

4.6 Multiplicity Adjustment

Not applicable.

4.7 Handling of Dropouts and Missing Data

Subjects who discontinued prematurely will not be replaced. Missing dates or partially missing dates will be imputed conservatively for concomitant medications and AEs to ensure that an AE is considered treatment emergent and has the longest possible duration.

For the efficacy endpoints, the missing data will not be imputed.

5 SUBJECT DISPOSITION

The following disposition information will be summarized by age group at enrollment and all ages combined:

- Number of subjects enrolled in Part 1 and continue to Part 2
- Number of subjects who completed the study and each Part
- Number of subjects who discontinued the study and each Part
- Number of subjects in each analysis population

The number and percentage of subjects who terminated early will be summarized separately by age group at enrollment, and all ages combined. The reasons for early termination will also be summarized and provided in the same summary table. A subject listing of completion and early termination will be provided.

6 PROTOCOL DEVIATIONS

A major protocol deviation is defined as a departure from the approved study protocol that may impact the rights, safety, or welfare of the subjects or the integrity of the data, which will be summarized by deviation category.

A minor or administrative protocol deviation is defined as a departure from the approved study protocol that has minimum or no impact on the rights, safety, or welfare of the subjects or the integrity of the data. Minor protocol deviations will also be summarized by deviation category. For details regarding the classification of the deviations, please refer to the Protocol Deviation Plan.

The number and percentage of subjects with major protocol deviations and type of deviation will be summarized by age group at enrollment, and all ages combined in the Enrolled population. A summary of the inclusion/exclusion criteria deviations, including exemptions, and protocol deviations will be provided. In addition, a list of subjects enrolled who did not meet the entrance eligibility criteria and subjects who did not adhere to the protocol after entry will be provided.

7 DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Subject demographics (age, sex, ethnicity, and race) and baseline characteristics (weight, height, head circumference, blood Phe concentration, prescribed dietary Phe, and blood tryptophan and tyrosine concentration) will be summarized separately by age group at enrollment and all ages combined. A separate table will summarize the neurocognitive testing (Bayley-III, WPPSI-III and WISC-IV) at Month 2 in Part 2.

Age, sex, ethnicity, and race will be summarized using the number and percentage of subjects with a particular attribute. The denominators for percentages will be based on the number of subjects in each age group. Weight, height, head circumference, blood Phe concentration, blood tryptophan and tyrosine concentration, and neurocognitive tests will be summarized using descriptive statistics including n, mean, SD, median, min, max, 25th percentile, 75th percentile. In addition, the percentile and z-scores of weight, height, and head circumference based on CDC Clinical Growth Charts (WHO Growth Charts for subjects less than 2 years of age) will be summarized using descriptive statistics.

A subject listing of demographics and baseline characteristics will also be provided.

8 MEDICAL HISTORY

Each subject's chronic diseases, disorders, and surgeries in the past will be collected as general medical history. General medical history will be summarized using MedDRA by system organ class (SOC) and medical history by descending order of frequency for all subjects in the Enrolled and Safety populations. A by-subject listing of each subject's medical history collected at the Screening Visit will also be provided.

9 PRIOR AND CONCOMITANT MEDICATIONS/PROCEDURES

Prior and concomitant medications will be summarized across all subjects and by age group in the Enrolled and Safety populations. For analysis purposes, the following definitions will be used to determine prior and concomitant medications:

- Prior medications: any medications taken within 30 days prior to screening and prior to the date of initial study drug administration in Study PKU-015.
- Concomitant medications: any medications initially taken on or after the date of initial study drug administration in Part 1.
- Prior and concomitant medications: any medications taken both prior to the date of initial study drug administration and on or after the date of initial study drug administration in Part 1will be reported both as prior and concomitant medications. Therefore, any medications taken prior to the first date of Part 1where the stop date is reported as "continuing" or missing will be considered both as prior and concomitant medications.

In the event the start date of a medication is completely missing, the start date will be imputed as the first dose date if the stop date is in the same year or later than the first dose date.

In the event the start date of a medication is partial, the following imputation rules will be applied:

- If only day is missing and year and month is after the first dose year and month, then the start date will be imputed as the first day of the month. If only day is missing and the year and month is prior to the first dose year and month, then the start date will be imputed as the last day of the month. If month is the same as the month of first dose of study drug then the start date will be imputed as the first dose date.
- If only year is non-missing, then the start date will be imputed as the first day of the year if the year is after the year of first dose. If year is the same as the year of first dose of study drug then the start date will be imputed as the first dose date. If year is before the year of first dose of study drug then the start date will be imputed to the first day of July of the year.

In the event the stop date of a medication is partial, the following imputation rules will be applied:

- If only day is missing, then the end date will be imputed as the last day of the month.
- If only year is non-missing, then the end date will be imputed as the last day of the year.

All medications will be coded using the current version of the World Health Organization Drug (WHO Drug) dictionary at program termination or annual reporting. Prior and concomitant medication use will be separately summarized by Anatomical Therapeutic

Chemical (ATC) medication class (Level 4) and preferred name (ie, generic medication name). Concomitant medication use will also be summarized. A subject reporting the same medication more than once will be counted once when calculating the number and percentage of subjects who received that medication. A by subject listing of each subject's prior and concomitant medications will be provided.

10 DIETARY MANAGEMENT

Occurrence of dietary modification and reason for modification will be listed.

11 COMPLIANCE

Treatment compliance will be summarized by age group at enrollment and all ages combined. In addition, the treatment compliance will be summarized in 3 categories: <80%, 80% - 120% (inclusive), and > 120%. The compliance to study drug will be measured in the following parameters:

- Treatment compliance based on prescribed dose taken

$$\text{Formula: } \frac{\text{Total Kuvan Intake}}{\text{Total Kuvan Prescribed}} \times 100\%$$

The total amount of Kuvan intake (measured in mg) will be calculated by multiplying the actual dose of daily Kuvan taken by the duration of the dosing period and then summing across all dosing periods. The total amount of Kuvan prescribed will be determined by multiplying the prescribed daily Kuvan by the duration of the dosing period and then summing across all dosing periods. The duration of a dosing period is defined as the subject's medication stop date minus the medication start date plus 1 for each dosing period.

- The percentage of days on correct dosage

$$\text{Formula: } \frac{\text{Number of Days with Correct Dose Intake}}{\text{Total Days of All Dose Periods}} \times 100\%$$

The numerator is the total number of days the subject took the same amount of study drug as the prescribed dose. The total days of all dosing periods will be calculated by summing the durations of all dosing periods.

- The percentage of days on drug

$$\text{Formula: } \frac{\text{Number of Days with Kuvan Intake}}{\text{Total Days of All Dose Periods}} \times 100\%$$

The numerator is the number of days the subject took Kuvan, regardless of the amount taken. The total days of all dosing periods will be calculated by summing the durations of all dosing periods.

- Actual daily dose taken (mg/kg/day)

The parameter will be calculated as the mean of daily dose from all dosing periods. Daily dose from each dosing period will be calculated by sum of total amount of Kuvan intake in mg, divided by dosing weight, and the duration of the dosing period.

A subject listing of treatment compliance will be provided.

12 EXTENT OF EXPOSURE TO STUDY DRUG

The duration over which subjects took Kuvan will be derived from the study drug exposure Case Report Form (CRF). The durations of exposure are the sum of days that the subject took Kuvan. Descriptive statistics for total duration will be provided by age group at Screening, study part, and overall. The number of dose changes and the reasons for the changes will be summarized as well.

A by-subject listing of each subject's extent of exposure and dose changes will be provided.

13 EFFICACY EVALUATIONS

Unless specified otherwise efficacy analyses will be based on the efficacy population. The efficacy results will be presented by age group at Screening and by all ages combined, except for the neurocognitive assessments.

13.1 Primary Efficacy Endpoint

The primary efficacy variable is the FSIQ score measured by the WPPSI-III or WISC-IV tests. The primary efficacy endpoint is the change in the FSIQ score over a 2 year period. A random coefficient model is used to calculate the slope (per year) of FSIQ over the entire study period and for each subject. Factors in the model will include visit and testing sequence, with change in FSIQ score as the dependent variable. Random terms include both intercept and visit. The model will use an unconstructed covariance matrix unless another covariance structure provides a better model fit. Subjects with less than one year of treatment duration are excluded from the analysis. Once the slope and 95% confidence interval are derived, the slope and lower limit of the confidence interval are multiplied by 2 to determine the average change in FSIQ and its lower limit over a 2-year window. The treatment will be considered successful if the lower 95% confidence limit of the mean change excludes a decline of greater than 5 points over a 2-year window.

All available data will be included in the analysis and the analysis will be stratified by testing sequence.

- Sequence 1: WPPSI-III (first or second form) at all time points
- Sequence 2: WPPSI-III (first or second form) at Baseline and (WPPSI-III or) WISC-IV at all post-baseline time points
- Sequence 3: WISC-IV at all time points

13.1.1 Supportive Analysis of the Primary Endpoint

A supportive analysis to the primary endpoint will be the mean difference between the last two available FSIQ assessments that were two years apart. This analysis was conducted to provide additional data on the 2-year analysis window when the subject was oldest, as the reliability of IQ testing in children increases over time. A 95% confidence interval will be calculated using a paired t-test. The result of a lower 95% confidence limit of the mean

difference excluding a decline of greater than 5 points is deemed favorable and supports a positive primary analysis result.

13.2 Secondary Endpoints

The secondary efficacy variables are:

- Growth measurements (height, weight, head circumference z-scores)
- Neurocognitive testing results (Bayley-III) in children age 0 to less than 30 months old.

A subject listing of Bayley-III results and growth parameters (height, weight, and head circumference assessments and corresponding percentile and z-score) will be provided..

13.2.1 Growth Measurements

The change in height, weight, and head circumference from baseline will be assessed at each scheduled visit. In addition, height, weight, and head circumference will be converted to z-score based on CDC Clinical Growth Charts (WHO Growth Charts for subjects less than 24 months) and the percentile, z-scores, and change in z-score from baseline of these growth variables will be summarized for each age group and all ages combined at each schedule time point. Growth charts using z-scores will be provided in each scheduled time point.

13.2.2 Neurocognitive Testing (Bayley-III)

The change in each composite score from baseline will be assessed at each scheduled visit. The three composite measures are cognitive, language, and motor. Only subjects less than 30 months of age are considered.

13.3 Exploratory analyses

Exploratory variables include:

- IQ scores
- Blood Phe concentration
- Relationship between neurocognitive testing and blood Phe
- Prescribed dietary Phe

A subject listing of IQ scores, neurocognitive testing results, blood Phe, and prescribed dietary Phe will be provided.

13.3.1 IQ Scores

An exploratory analysis to assess the effect of blood Phe control with Kuvan on IQ over a 7 year observation period will examine the stability from baseline in the FSIQ score. Baseline for FSIQ score is defined as the first assessment in FSIQ score after determination of Kuvan responder status. The analysis will be performed using a repeated measures model with factors visit and testing sequence, with change in FSIQ score as the dependent variable. The model will apply an unconstructed covariance matrix unless another covariance structure provides a better model fit. 95% Confidence intervals will be reported annually with emphasis at Year 7. A figure of LS-means and the 95% CI will also be displayed.

Specific subsets are summarized to further examine the robustness of the above Year 7 results. The following subsets will be used:

- Subjects enrolled in study due to an exemption are excluded
- Subjects with a Baseline FSIQ score at the recommended visit (Month 2)
- Subjects assigned to Sequence 2 with Baseline as the first assessment in FSIQ score after determination of Kuvan responder status
- Subjects assigned to Sequence 2 with Baseline as the first assessment in FSIQ score at the recommended visit (Month 2)

All subset summaries include a repeated measure analysis similar to the original analysis. Sequence is removed from the analyses where the subset is based on sequence.

Furthermore, the mean change from baseline in FSIQ score at Year 7 will be summarized descriptively by sequence and overall. A paired t-test will be performed across all subjects with a 95% CI and p-value reported at Year 7. An additional subset examining subjects without a Baseline FSIQ score at the recommended visit (Month 2) is presented as well.

In addition, descriptive statistics of each observed composite and subtest IQ score and its respective change in IQ score from Baseline will also be presented by sequence and overall and listed at each time point. The change from Baseline of each composite score will be plotted as well.

Proportion of subjects with a specified change from Baseline in each composite and subtest IQ will also be summarized at each time point. The specific change is as follows:

- FSIQ: absolute change greater than 7
- Verbal IQ: absolute change greater than 10
- Performance IQ: absolute change greater than 10
- Each subtest: absolute change greater than 3

13.3.2 Blood Phe Concentration

The observed value and change from Baseline in blood Phe will be summarized at each scheduled visit and across all visits. In addition, the average proportion of Phe values between the target range of 120 and 360 across each subject will be presented.

Another measurement of interest is the Index of Dietary Control (IDC), defined as half-year Phe level medians averaged over the first 6 months, over the first year, and annually thereafter. Descriptive statistics of the IDC will be produced.

Figures displaying mean change in Phe from baseline and IDC at each time point will be produced.

13.3.3 Relationship between Neurocognitive Scores, Blood Phe, and Prescribed Dietary Phe

The relationship between components of the WPPSI-III or WISC-IV and blood Phe will be examined. Specifically,

1) Observed values in FSIQ and observed values in blood Phe at end of study
2) Observed values in other components and observed values in blood Phe at end of study
3) Change in FSIQ and change in blood Phe at end of study
4) Change in other components and change in blood Phe at end of study
5) Observed values in FSIQ and observed values in prescribed dietary Phe at end of study
6) Observed values in other components and observed values in prescribed dietary Phe at end of study
7) Change in FSIQ and change in prescribed dietary Phe at end of study
8) Change in other components and change in prescribed dietary Phe at end of study

9) Change in FSIQ at end of study and SD of blood Phe over the study period
10) Change in other components at end of study and SD of blood Phe over the study period

For the above analyses, each subject's final assessment should occur within the same visit. Because each subject will have a different duration, study duration will be summarized as well. For items 9 and 10 above, these analyses will be repeated for subjects with assessments at Month 60 and at least one additional assessment after Month 60. In this analysis, the changes are calculated from Month 60, not Baseline.

Scatterplots will be presented and correlation calculated for each analysis.

13.3.4 Prescribed Dietary Phe

The observed value and change from Baseline in prescribed dietary Phe will be assessed at each scheduled visit.

In addition, the relationship between the change in z-score weight and height will be examined alongside the change in prescribed dietary Phe. The final assessment of weight, height, and dietary Phe should occur within the same visit. Scatterplots will be presented and correlation calculated for each analysis.

14 SAFETY EVALUATIONS

Safety will be assessed by examining the incidence, severity grade, and relationship to study drug of all treatment-emergent adverse events (TEAEs) reported during the study period. In addition, clinical laboratory results, vital signs, ECG, and physical exam will be assessed. No formal statistical testing will be performed for the safety analyses; only descriptive statistics will be provided. Safety summaries will be presented for the Safety population using both group presentations as described in Section 4.2 unless noted otherwise.

14.1 Adverse Events

14.1.1 All Adverse Events

Adverse events will be coded in accordance with the most current version of MedDRA at program termination or annual reporting. Only treatment-emergent adverse events (TEAEs) occurring and reported during the study will be included in the adverse event summaries. A TEAE is defined as any AE that newly appeared or worsened in severity following initiation of the study drug administration. If the onset date of an AE is missing, the AE will be considered treatment-emergent. The incidence of all treatment-emergent AEs will be summarized by system organ class (SOC), preferred term (PT), relationship to study drug and CTCAE grades. For those AEs that occurred more than once, regardless of study part, the maximum severity will be used to summarize the AEs by severity.

If the onset date or end date of an adverse event is partial, the same imputation rules described in section 9 will be applied.

The treatment-emergent AEs will be summarized in the following categories:

- All AEs by SOC and preferred term
- All AEs by preferred term in descending frequency order
- All AEs by SOC, preferred term, and NCI CTCAE grade
- AEs assessed by investigators as related to study drug by SOC and preferred term
- AEs assessed by investigators as related to study drug by SOC, preferred term and NCI CTCAE grade
- Serious AEs (SAEs) by SOC and preferred term
- All SAEs by preferred term in descending frequency order

- SAEs assessed by investigators as related to study drug by SOC and preferred term
- Death
- AEs leading to withdrawal from study by SOC and preferred term
- AEs leading to study drug discontinuation by SOC and preferred term

A summary of all AEs and adverse event listings will also be provided.

14.1.2 Adverse Events of Special Interest

Occurrences of AEoSIs are derived based on the search strategy below. AEoSIs will be summarized in a similar manner to general adverse events for any and each AEoSI and listed as well. Only summaries involving Part 2 subjects are examined.

In addition, the exposure-adjusted incidence rate for each AEoSI will be summarized. This rate will be calculated as the number of individuals with the event starting at first study drug administration, divided by the total number of patient-years at risk for the event, multiplied by 100. In calculating patient-years at risk for a given event, the follow-up time for a patient experiencing that event will conclude at the event start date. If the subject never experience the event, the subject will be censored at the discontinuation visit, death, or last follow-up visit, whichever is earlier. Mean cumulative function plots will also be produced for any AEoSI if the number of events for that specific AEoSI are sufficient.

The following adverse events are of interest:

TERM	STRATEGY
hypophenylalanenemia	MedDRA Preferred Terms(PTs): Phenylketonuria, Amino acid level decreased and Amino acid level abnormal
convulsions	Standardised MedDRA queries(SMQs): Convulsions
epigastric ulcer	Standardised MedDRA queries(SMQs): Gastrointestinal ulceration, Gastrointestinal perforation and Gastrointestinal haemorrhage
nephrotoxicity	Standardised MedDRA queries(SMQs) of Broad acute renal failure, Broad chronic kidney disease *or* MedDRA HLGT Term(s) of Nephropaties and Renal disorders (excl nephropathies)
behavioural changes	MedDRA HLGT Term(s): Behaviour and socialisation disturbances

TERM	STRATEGY
nephrolithiasis	MedDRA HLGT Term(s): Renal lithiasis
GERD	MedDRA Preferred Terms(PTs): Gastroesophageal reflux disease, Barrett's esophagus, and Gastrointestinal nonspecific inflammation and dysfunctional disorders
Hypersensitivity Reactions	Standardized MedDRA Queries(SMQs): Broad Hypersensitivity reaction
Anaphylactic reaction - Broad Algorithmic anaphylactic reaction SMQ	Standardized MedDRA Queries(SMQs): Broad algorithmic anaphylactic reaction
New onset anxiety disorders	MedDRA HLGT Term(s): Anxiety disorders and symptoms
Worsening of psychiatric disorders	MedDRA System Organ Class (SOC): Psychiatric disorders
Drug-drug interaction	MedDRA PTs: Phenylketonuria and Neurotransmitter level altered

14.2 Clinical Laboratory Tests

Standard descriptive statistics will be presented for clinical laboratory tests, including hematology, chemistry and urinalysis, at each scheduled visit and for change in laboratory values from baseline. The ratio of Phe and Tyrosine will be summarized descriptively at each scheduled visit as well.

Blood levels of Tyrosine, tryptophan, and albumin will be compared to age-related norms. The proportion of subjects who experience post-baseline laboratory tests result outside the age-related normal range will be presented for each measurement by the appropriate age category annually. Lab measurements will be averaged if more the one value is reported within the year.

The proportion of subjects who experienced at least one laboratory test result outside the normal ranges will be presented for each measurement. Shift tables will be created to summarize the change in laboratory value from baseline to worst post-baseline value based on normal range (low, normal, and high) or NCI CTCAE grading, where available. A subject listing of laboratory test results for hematology, chemistry, urinalysis, and other laboratory tests will be provided separately. Abnormal values and clinically significant abnormal values will be flagged in the listings.

14.3 Vital Signs

Standard descriptive statistics for temperature, respiratory rate, heart rate, and blood pressure (systolic and diastolic) will be presented at each scheduled visit. Additionally, change from baseline to each scheduled visit will be summarized by scheduled visit using descriptive statistics.

A subject listing of vital signs reported will be also provided.

14.4 Electrocardiogram

All ECG evaluations will be presented as data listings.

14.5 Physical Examination

All physical examination results will be presented as data listings.

15 SUMMARY OF CHANGES TO STUDY SAP

The primary differences between Version 1.0 and Version 2.0 of the statistical analysis plan are:

- Section 4.1 has been updated to reflect additional populations and updated definitions.
- Section 4.2 has been amended to include an additional grouping for safety analyses.
- Section 4.5 has been updated to clarify visit windows
- Section 13.1 has been revised to better reflect the analysis performed
- Section 13.1.1 originated
- Section 13.2.1 has been revised to better reflect the analysis performed
- Section 14 has been revised to better reflect the analysis performed

- Section 14.1.2 has been updated to modify definitions and add terms of interest

The primary differences between Version 2.0 and Version 3.0 of the statistical analysis plan are:

- Section 13 has been revised with to align the primary endpoint with the protocol, and to clarify supportive, secondary, and exploratory analyses. The analyses in Section 13.1 are original to this document.
- Section 14.1.2 has been updated to remove gastritis as an AE of special interest (AEoSI). Gastritis has been assessed as not an AEoSI and is not listed as such in any periodic safety report or the Risk Management Plan, and does not require reporting beyond regular pharmacovigilance surveillance.

16.1.9.2 Errata Table

BioMarin Pharmaceutical Inc.

ERRATA FORM

Protocol Number PKU-015
Date 08Jan2019

The following discrepancies were noted and per BioMarin, these data points will not be updated/changed in the database.

Errata:

Source	Inv #	Patient #	Table Name	Field Name	Explanation
CONMED	[redacted]	[redacted]	14.1.6.1	Dose Per Administration	Tamiflu should be recorded in EDC as 1 Caps BID for Flu from [redacted]2009 to [redacted]2009. However EDC currently indicates 10 Caps BID for Flu from [redacted]2009 to [redacted]2009.

[redacted] 18JAN2019

Study Statistician (Print Name) Date

[redacted]

Study Statistician Signature Date