Sobi 003-002 Extension Study

Statistical Analysis Plan

Protocol number:	SOBI003-002
Title:	**An open, single-arm, multicenter extension study to assess the safety, tolerability and efficacy of long-term SOBI003 treatment in pediatric MPS IIIA patients**
SAP Author:	[redacted]
SAP version:	1.0 Final Version, dated 13 September 2018

Signature Date

Author

[redacted]

Table of contents

Table of figures

Table of tables

1 Abbreviations and definition of terms

ADA	Anti-drug antibody
AE	Adverse event
AEq	Age-equivalent score
AUC	Area under curve
BSID-III	Bayleys Scales of Infant and Toddler Development®, third edition
CL	Clearance
CNS	Central nervous system
CSR	Clinical study report
DQ	Development quotient
FAS	Full analysis set
FIH	First time in human
HS	Heparan sulfate
IMP	Investigational medicinal product
KABC-II	Kaufman assessment battery for children, second edition
MPS IIIA	Mucopolysaccharidosis type IIIA
MAR	Missing at random
MRI	Magnetic resonance imaging
NH	Natural history
NVI	Nonverbal index
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SRC	Safety review committee
Sobi	Swedish Orphan Biovitrum
TEAE	Treatment emergent adverse event
PPS	Per Protocol Set
VABS-II	Vineland adaptive behavior scales, second edition

2 Introduction

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for Sobi protocol SOBI003-002 (An open, single-arm, multicenter extension study to assess the safety, tolerability and efficacy of long-term SOBI003 treatment in pediatric MPS IIIA patients).

This extension study is being completed to assess the efficacy and safety of SOBI003 for the treatment of MPS IIIA in pediatric patients.

The purpose of this SAP is to outline the planned analyses to be completed to support the Clinical Study Report (CSR) for protocol SOBI003-002. The planned analyses identified in this SAP will be included in regulatory submissions and/or future manuscripts. Also, exploratory analyses not necessarily identified in this SAP may be performed to support the clinical development program. Any post-hoc, or unplanned, analyses not identified in this SAP will be clearly identified in the respective CSR. This SAP also includes a description of upcoming Phase III analyses, see Section 17.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the FDA and International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH): Guidance on Statistical Principles in Clinical Trials. All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association and the Royal Statistical Society, for statistical practice.

3 Study objectives and endpoints

3.1 Primary objective

The primary objective is:

- To assess the long-term safety and tolerability of SOBI003

3.2 Key secondary objectives

- To evaluate the efficacy of SOBI003 on neurocognitive function, as compared to untreated MPS IIIA patients from a NH control.
- To evaluate the effect of SOBI003 on HS levels in CSF

3.3 Secondary objectives

The secondary objectives are:

1. To assess the effect of SOBI003 on adaptive behavior, as compared to untreated MPS IIIA patients from a NH control
2. To assess the effect of SOBI003 on HS levels in serum and urine

3. To assess the effect of SOBI003 on brain magnetic resonance imaging (MRI) abnormalities
4. To assess the effect of SOBI003 on liver and spleen volume
5. To assess the effect of SOBI003 on Quality of Life
6. To assess the effect of SOBI003 on language
7. To assess the effect of SOBI003 on motor function
8. To assess the effect of SOBI003 on sleep pattern
9. To assess the immunogenicity of SOBI003
10. To characterize steady-state pharmacokinetics (PK) of SOBI003 by the use of non-compartmental analysis (NCA)

3.4 Exploratory objectives

The exploratory objectives are:

1. To assess the effect of SOBI003 on adaptive behavior over time
2. To characterize the PK properties of SOBI003 following repeated administration by the use of population PK analysis
3. To evaluate the PK/PD relationship between SOBI003 concentrations in serum and effect of SOBI003 on HS levels in CSF, serum and urine by the use of population modelling analysis
4. To evaluate the PK/PD relationship between SOBI003 concentrations in serum and CSF, and the effect of SOBI003 on other biomarkers than HS and neuro-cognitive measures eg. DQ, as well as MRI data, by the use of population modelling

As local regulations permit and provided that additional separate caregiver consent is given, the exploratory objectives are also to:

- Collect and store serum and CSF samples to enable analyses of biomarkers with possible relation to safety, tolerability, immunogenicity, PK and PD of SOBI003, as identified in future

3.5 Study endpoints

3.5.1 Primary endpoint

The primary endpoint to evaluate the safety and tolerability of SOBI003 is:

- Treatment Emergent Adverse Events (TEAEs) and Serious Adverse Events (SAEs)

3.5.2 Secondary endpoints related to the primary objective

The secondary endpoints to evaluate the safety and tolerability of SOBI003 are:

- Vital signs (blood pressure, heart rate, body temperature, respiratory rate and pulse oximetry)
- Laboratory safety variables (hematology, coagulation, clinical chemistry and urine analysis)

3.5.3 Key secondary endpoints

Study efficacy evaluation will primarily be based on:

- Change from baseline at Week 104 in Development Quotient (DQ) as assessed by the Bayley Scales of Infant and Toddler Development®, third edition (BSID-III) cognitive subtest or the Kaufman Assessment Battery for Children, Second edition (KABC-II)
- Change from baseline at Week 104 in CSF HS

Additional key secondary endpoints are:

- Change from baseline at Week 52 in Development Quotient (DQ) as assessed by the Bayley Scales of Infant and Toddler Development®, third edition (BSID-III) cognitive subtest or the Kaufman Assessment Battery for Children, Second edition (KABC-II)
- Change from baseline at Weeks 52 and 104 in Age Equivalent (AEq) as assessed by the Bayley Scales of Infant and Toddler Development®, third edition (BSID-III) cognitive subtest or the Kaufman Assessment Battery for Children, Second edition (KABC-II)
- Change from baseline at Week 52 in CSF HS

3.5.4 Secondary endpoints

The endpoints relating to the 1nd secondary objective are:

- Change from baseline at Weeks 52 and 104 in adaptive behavior age-equivalence score (AEq) as assessed by Vineland Adaptive Behavior Scales, Second edition (VABS-II)
- Change from baseline at Weeks 52 and 104 in adaptive behavior composite score as assessed by Vineland Adaptive Behavior Scales, Second edition (VABS-II)

The endpoints relating to the 2rd secondary objective are:

- Change from baseline at Weeks 52 and 104 in serum HS
- Change from baseline at Weeks 52 and 104 in urine HS

The endpoints relating to the 3th secondary objective are:

- Change from baseline at Weeks 52 and 104 in gray matter volume as assessed by brain volumetric MRI
- Change from baseline at Weeks 52 and 104 in compound ventricular volume as assessed by brain volumetric MRI
- Change from baseline at Weeks 52 and 104 in fractional anisotropy (FA) and mean diffusivity (MD) of corpus callosum as assessed by diffusion tensor imaging MRI

- Change from baseline at Weeks 52 and 104 in FA and MD of cerebral white matter as assessed by diffusion tensor imaging MRI
- Change from baseline at Weeks 52 and 104 in cerebral white matter as assessed by susceptibility weighting imaging (SWI) MRI
- Change from baseline at Weeks 52 and 104 in basal ganglia as assessed by SWI MRI

The endpoints relating to the 4th secondary objective are:

- Change from baseline at Weeks 52 and 104 in liver volume as assessed by abdominal MRI
- Change from baseline at Weeks 52 and 104 in spleen volume as assessed by abdominal MRI

The endpoints relating to the 5th secondary objective are:

- Change from baseline at Weeks 52 and 104 in Pediatric Quality of Life inventory (PedsQL) total score

The endpoints relating to the 6th secondary objective are:

- Change from baseline at Weeks 52 and 104 in expressive language AEq score as assessed by the BSID-III language subtests or the KABC-II Expressive Vocabulary subtest, and the VABS-II communication domain
- Change from baseline at Weeks 52 and 104 in receptive language AEq score as assessed by the BSID-III language subtests or the KABC-II Expressive Vocabulary subtest, and the VABS-II communication domain

The endpoints relating to the 7th secondary objective are:

- Change from baseline at Weeks 52 and 104 in gross motor function AEq score as assessed by the BSID-III motor subtests and the VABS II
- Change from baseline at Weeks 52 and 104 in fine motor function AEq score as assessed by the BSID-III motor subtest and the VABS-II

The endpoints relating to the 8th secondary objective are:

- Change from baseline at Weeks 52 and 104 in sleep pattern as assessed by the Children´s Sleep Habits Questionnaire (CSHQ) score
- Change from baseline at Weeks 52 and 104 in sleep-related variables as assesses by actigraphy (including total sleep time, total day- and night time sleep duration, sleep latency, sleep efficiency, number of nocturnal awakenings, and wake after sleep onset)

The endpoints relating to the 9th secondary objective are:

- Occurrence of Anti-drug antibody (ADA) against SOBI003 in serum at Weeks 39, 52, 78 and 104, and adjacent to dose adjustments (seroconversion rate, time to seroconversion, transient/persistent). For patients with confirmed ADA positive serum samples, the following additional endpoints apply; ADA titers and IgG subclasses in serum, and presence of neutralizing antibodies (NAb) in serum.
- Occurrence of ADAs against SOBI003 in CSF at Weeks 39, 52, 78 and 104 (conversion rate, time to occurrence, transient/persistent). For patients with confirmed ADA positive

CSF samples, the following additional endpoints apply: ADA titers and presence of NAb in CSF.

The endpoints relating to the 10th secondary objective are:

- Serum SOBI003 PK parameters at Weeks 38, 52, 78 and 104, and adjacent to dose adjustments; $t_{End\ of\ inf}$, $C_{End\ of\ inf}$, $C_{Pre\text{-}dose}$, CL, $AUC_{0\text{-}168h}$
- SOBI003 concentrations in CSF pre-dose at Weeks 52 and 104.

3.5.5 Exploratory endpoints

The endpoints related to the 1st exploratory objective are:

- Change from baseline at Weeks 52 and 104 in adaptive behavior composite score as assessed by Vineland Adaptive Behavior Scales, Second edition (VABS-II)

The endpoints related to the 2nd exploratory objective are:

- Population PK model parameter estimates and associated covariates describing intra- and inter-individual variability in respective parameter estimate. The results of these analyses will not be included in the Clinical Study Report for this study, but reported separately.

The endpoints related to the 3rd exploratory objective are:

- Population PK/PD model parameter estimates and associated covariates describing intra- and inter-individual variability in respective parameter estimate. The results of these analyses will not be included in the Clinical Study Report for this study, but reported separately.

4 Study methods

4.1 Overall study design and plan

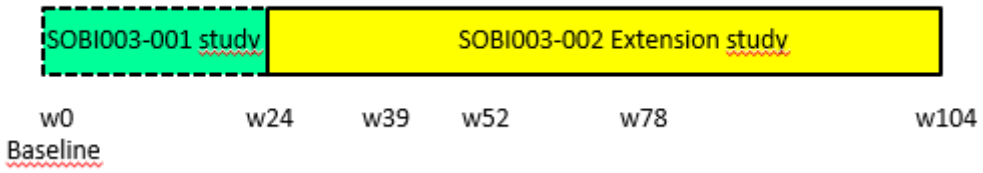
This is an open, single-arm, multicenter extension study to assess the long-term safety, tolerability and efficacy of SOBI003 in pediatric MPS IIIA patients who previously participated in the (FIH) SOBI003-001 study. Patients completing the FIH study will be offered to participate in the extension study. Please refer to Figure 1 for an overview of the study design. A published natural history study (referred to as the NH study in the present document) is used as external control (**Error! Reference source not found.**). Neurocognitive and adaptive behavior data assessed in the extension study at Week 52 and Week 104 following start of treatment will be compared to available natural history data within the target study population. The natural history population for comparison consists of 14 patients with a rapid progressive MPS IIIA disease and a baseline age of 12 to 72 months.
As the FIH study has a sequential, ascending, multiple-dose design, the first patients will enter the extension study (SOBI003-002) before finalization of the FIH study. When entering the extension study, these patients will receive the highest dose that has been declared safe by the

SRC in the ongoing FIH study. Patients enrolled in Cohort 1 of FIH may thus be switched to the dose applied in Cohort 2 or Cohort 3, depending on safety clearance by the SRC. If a patient is initially switched to the dose applied in Cohort 2, the patient may subsequently later be switched to the dose applied in Cohort 3. Depending on recruitment rate in the FIH study, a patient enrolled in Cohort 1 may be directly switched to the dose applied in Cohort 3. Upon completion of the FIH study, an analysis aimed at selecting the dose for forthcoming studies will take place. Once the dose has been selected, this dose will be applied to all patients enrolled in the extension study. The Baseline in SOBI003-001 will be used as the baseline for SOBI003-002 (Figure 1).

SOBI003 is administered as weekly i.v. infusions over a period of time of 1 to 4 hours. Prior to initiation of each infusion, the patients are pretreated with a single dose of non-sedative antihistamine. If infusion-related reactions occur, then the infusion duration may be expanded up to 24 hours and supportive medication may be administered, at the discretion of the investigator.

The long-term safety, tolerability and immunogenicity of SOBI003 will be assessed throughout the extension study whereas measurements of clinical efficacy and disease-related biomarkers will be performed at Weeks 52 and 104. The potential of SOBI003 to improve the neurocognitive function will be compared to published data from the NH study.

Figure 1 Overview for study design for study SOBI003-002



4.2 Selection of study population

The primary aim is to select a study population as close as possible to the NH study.

4.2.1 Inclusion criteria

A patient must fulfill the following criteria in order to be included in the study:

1. Completion of study FIH
2. Informed consent obtained from the patient's legally authorized representative(s)

4.2.2 Exclusion criteria

The presence of the following excludes a patient from inclusion in the study:

1. If, in the opinion of the investigator, there are patient specific safety concerns that contraindicates further treatment with SOBI003

4.3 Method of treatment assignment

4.3.1 Control group

The control group used for the statistical analysis of the DQ endpoint is formed by the 14 patients in the Shapiro Natural History Study with a rapid progressive disease who have a baseline age of 12 to 72 months (Shapiro et al 2016).

4.3.2 SOBI003

All patients in this extension study will receive SOBI003 until 104 weeks after start of treatment unless withdrawn.

5 Sequence of Planned analyses

5.1 Interim Analyses

There is no planned interim analysis for this extension study.

5.2 Analyses and reporting

All final, planned, analyses identified in the protocol and in this SAP will be performed only after the last patient has completed the study.

Any post-hoc analyses included in the CSR, which were not identified in this SAP, will be clearly identified as such in the relevant section of the CSR.

For each patient in the study, all the associated safety and efficacy data in the SOBI003-001 study will be pooled and presented together with the SOBI003-002 study. Thus the SOIB003-001 study and the SOBI003-002 study are considered as a single, unified entity.

6 Estimands

There will be a de facto estimand in the study, which is the primary estimand for the evaluation of efficacy. The estimand is the change in

- cognitive decline in all the patients from the population captured by the inclusion and exclusion criteria (see Section 4.2.1 and Section 4.2.2) and the part of the NH population with a rapid progressive disease having a baseline age of 12 to 72 months, irrespectively of other medications used and adherence to treatment, and estimated by a repeated measures model of the change from baseline in DQ as described in Section 14.1.

- decline in HS levels in CSF in all the patients from the population captured by the inclusion and exclusion criteria (see Section 4.2.1 and Section 4.2.2) irrespectively of other medications used and adherence to treatment, and estimated by a repeated measures model of the change from baseline in HS as described in Section 14.1.

To address foreseeable inter-current events, such as drop-outs, all missing values will be imputed (see Section 14.1) in accordance with the treatment policy principle. A sensitivity analysis will be performed where no imputations are performed, but all values are used: If a patient has a missing value in any assessment, then the same patient's non-missing scores will be used in the repeated measures model. The repeated measures model (see Section 14.1) handles missing values based on the missing at random (MAR) assumption. One further sensitivity analysis for DQ will address this assumption based on zero imputations (see Section 14.1.1), the rational being that the patients not assessed are those expected to have a severe DQ decline. Other sensitivity analyses for this purpose will be tipping point analyses, see Section 14.1.1.

Furthermore, a principal strata analysis will be performed where only patients without protocol violations will be analyzed according (see Section14.1.2). In particular patients who did not comply with the dosing scheme but continued the study will be excluded from the analysis. Thus the principal strata will be approximated by the clinical per protocol subset of FAS.

7 Sample size determination

7.1 Change from baseline in DQ score and in HS levels in CSF at 104 weeks

All patients completing the FIH study will be offered to continue into this extension study and the number of patients in the extension study is thus not based on statistical grounds.

However, the 14 patients in the natural history study that will be used as an external control group had a mean baseline DQ of 58.9 points and declined with a mean of 30.4 (SD=12.1) DQ points during 104 weeks. Therefore, assuming very high likelihood of showing significant decrease in HS (with baseline mean of 5.2, std 1.8, correlation of 0.1 and a decrease of 80 % the power is 99 %), with 9 patients treated with SOBI003, the power to demonstrate an improvement in cognitive function is at least 80 % if the decline is reduced from 30.4 to 9.8 DQ points (SD=13.0) and the reduction is homogeneous across age.

8 Analysis sets

The following analysis sets will be used in the statistical analyses and exploration of data.

8.1 Safety analysis set

The Safety analysis set will consist of all patients in the study. For each patient in the study, all the associated safety and efficacy data in the SOBI003-001 study will be pooled and presented together with the SOBI003-002 study. The safety set will be used for the safety presentations.

8.2 Full-analysis set

The full-analysis set (FAS) is the efficacy analysis set which will consist of all patients enrolled in this extension study, and all patients in the NH study with a rapid progressive disease having a baseline age of 12-72 months (in the DQ analyses).

All patients enrolled in this extension study and the FIH study will be treated as having received active treatment and patients from the NH study will be treated as having received no treatment. For each patient in the study, all the associated safety and efficacy data in the SOBI003-001 study will be pooled and presented together with the SOBI003-002 study.

8.3 PK analysis set

The PK analysis set will consist of those patients in the safety analysis set who have at least one SOBI003 serum concentration value, without any protocol deviation jeopardizing the PK evaluation. For each patient in the study, all the associated safety and efficacy data in the SOBI003-001 study will be pooled and presented together with the SOBI003-002 study.

The PK analysis set will be used for the PK analyses.

8.4 Immunogenicity analysis set

The immunogenicity set will consist of those patients in the safety set who have sufficient blood samples taken for ADA testing at baseline (screening visit) and at least one post-dose time-point.

For each patient in the study, all the associated safety and efficacy data in the SOBI003-001 study will be pooled and presented together with the SOBI003-002 study.

The immunogenicity set will be used for the immunogenicity analyses.

9 General issues for statistical analysis

All statistical tests will be two-sided and performed using a 5 % significance level, if not stated otherwise. Results will be presented as the estimated mean value for each treatment group, the estimated difference between groups, the associated 95 % two-sided confidence interval and P-value. P-values from statistical analyses will be presented to three decimal places with values below 0.001 displayed as <0.001 and confidence intervals will be presented to one more decimal places than the raw data.

Continuous data will be summarized using descriptive statistics: n, mean, standard deviation (SD), standard error of the mean (SE), median, minimum and maximum, unless otherwise indicated. Minimum and maximum will be presented to the same number of decimal places as the raw data and mean, standard deviation, standard error of the mean and median will be presented to one more decimal place than the raw data.

Categorical data will be summarized using counts and percentages. Percentages will be suppressed when the count is zero, however the category will still be displayed. The denominator for all percentages will be the number of patients within the treatment group for the population of interest, unless otherwise indicated. Percentages will be presented to one decimal place.

All data will be tabulated in individual patient data listings.

Statistical analyses will be performed using SAS software (SAS Institute Inc, Cary, North Carolina, United States).

9.1 Handling of missing data

All missing DQ values and HS in CSF values will be imputed. The procedure is fully described in Section 14.1. A sensitivity analysis will be performed where no imputations are performed, but all values are used: If a patient has a missing DQ score in any assessment, then the same patient's non-missing scores will be used in the repeated measures model. The repeated measures model (see Section 14.1) handles missing values based on the MAR assumption. One further sensitivity analysis will address this assumption based on zero imputations (see Section 14.1.1), the rational being that the patients not assessed are those expected to have a severe DQ decline. Another sensitivity analysis for this purpose will be a tipping point analysis, see Section 14.1.1.

If an adverse event (AE) has a completely missing onset date, then the AE will be considered a TEAE, unless the (partial) stop date indicates differently. A medication with a completely missing start date is considered a prior medication. A medication with a completely missing stop date is considered a concomitant medication. If an AE or a medication has a partial missing start or stop date, the following rules will be used to impute the date. The imputed date will be used to determine whether it is a TEAE for adverse event, or a prior or concomitant medication.

Table 1 Partial date derivation

Partial Missing Start or Stop Date	Derived Start Date	Derived Stop Date
Missing month and day, and the year is present	January 1st of that year or first dose date if the year is the same as the year of first dose date	December 31st of that year

Partial Missing Start or Stop Date	Derived Start Date	Derived Stop Date
Missing day, but year and month are present	First day of that month or first dose date if the year and month are the same as the year and month of first dose date	Last day of that month
Missing month, but year and day are present	Missing month derived as January or same as first dose month if the year is the same as the year of first dose	Missing month derived as December

9.2 Multicenter studies

In order to examine the uniformity of the treatment effect across centers, additional statistical analyses will be performed on the DQ endpoint, consisting of descriptive, summary statistics presented by center in the treatment group.

9.3 Derived and computed variables

Scores (AEq and DQ) for neurocognitive development

Neurocognition is assessed at Weeks 52 and 104, and during Early Termination, by use of the neurocognition domain of the BSID-III and/or the KABC-II Nonverbal Index (NVI).

The selection of neurocognitive assessment method is done at the baseline assessment of the FIH study (SOBI003-001) based on the algorithm presented in **Error! Reference source not found.**. The neurocognitive test(s) applied in the FIH study should also be used in the extension study.

Table 2 Selection of neurocognitive assessment method

Chronological age at the assessment time point	Development age-equivalence as established by VABS-II adaptive behavior composite	Neurocognitive test to apply
<42 months	Any age-equivalence	BSID-III
≥42 months	<36 months	BSID-III
≥42 months	36 to 42 months	BSID-III + KABC-II*
≥42 months	>42 months	KABC-II

* The BSID-III should be administered first, and there should be at least 2 days until the KABC-II assessment in order to reduce KABC-II outcome bias of patient tiredness.

The neurocognitive developmental age (months) is derived on the basis of the BSID-III cognitive total raw score and BSID-III age-normative data.

The neurocognitive developmental age (months) is derived from the mean AEq scores on the NVI of the KABC-II, where an AEq scores is the average score for a particular age.

Development Quotient

DQ is calculated as the AEQ for cognitive development divided by the chronological age of the child multiplied with 100, i.e.

$$DQ = 100*AEQ / Age$$

For patients completing an assessment of both BSID-III and KABC-II, the BSID-III score will be used as a basis for calculating the DQ of neurocognitive development at baseline. A calibration of overlapping AEq scores between BDID-III and KABC-II will be done to assess if a patient outgrows the BSID-III test during the duration of the study. If confirmed, assessments of KABC-II will form the basis for calculating DQ of neurocognitive development for remaining study assessments.

10 Patient disposition

The following patient disposition details will be summarized and will include the number and percentage of patients:

- Included in the each study population (Safety analysis set, efficacy analysis set, immunogenicity analysis set, PK analysis set and PD analysis set)
- Completed the study
- Prematurely withdrew (including a breakdown of the primary reasons for withdrawal)

These data will be listed. Individual reasons for withdrawal and protocol deviations will be presented in listings for all randomised patients.

11 Demographics and baseline characteristics

11.1 Demographics

Individual patient demographics (age, gender and race) and body measurement data (height, weight, body mass index and head circumference) at study start will be listed.

Age will be counted in months. Height will be measured in centimetres and weight in kilograms. BMI (kg/m^2) will be calculated from the height and weight.

These demographic characteristics and body measurements will be summarized for the total safety analysis set and the pivotal sub group.

11.2 Baseline characteristics

Other baseline measurements that will be listed include DQ, AEQ and genotype. The listing will include the NH study as well.

11.3 Medical history

Medical and surgical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) (version 16.1 or higher). All medical history events will be listed by system organ class (SOC) and preferred term. Surgical history will be presented in a separate listing.

12 Prior and concomitant medication

All prior and concomitant medication will be coded using the latest version of World Health Organization drug dictionary (WHO Drug) version June 1 2013.

Prior medications are defined as those for which the end date is prior to Day 1. Concomitant medications are defined as those which are indicated as 'Ongoing' on Day 1, those which start on or after Day 1 and those which start prior to Day 1 and have an end date on or after Day 1.

If medication dates are incomplete and it is not clear whether the medication was concomitant, it will be assumed to be concomitant.

The number and percentage of patients who took any concomitant medications will be presented by medication class and standardized medication names sorted alphabetically.

All prior and concomitant medications will be listed by reported name, medication class, standardized medication name, indication, dose, dose unit, frequency, route of administration, start date and end date or 'ongoing' flag and duration of use. Concomitant medications will be flagged.

13 Treatment compliance

There will be no measurement of, or analysis of, treatment compliance, nor does it affect population assignment (patients must be exposed to IMP to qualify for the analysis sets).

14 Efficacy analyses

14.1 The DQ endpoint

One of the key secondary objectives is to assess the effect of SOBI003 on neurocognition, as compared to untreated MPS IIIA patients from a NH control. The key endpoint related to this

objective is the change from baseline in DQ of neurocognitive development at 104 weeks of treatment, and will be estimated by a model based contrast.

To address the efficacy objective, a repeated measure analysis will be utilized where DQ is the response variable and time, baseline age, time*baseline age and treatment*time are fixed effects and the dependency between observations within the same patient is represented by an unstructured variance-covariance matrix.

The time window is two weeks, i.e., if the assessment is more than two weeks too early or more than two weeks too late, then the DQ is set to missing for that assessment.

In the NH study, a simple linear regression model was used, i.e. with DQ as the response variable and the age of the child as dependent variable (Shapiro et al 2016). As the aim in this clinical development program is to demonstrate that treatment with SOBI003 is efficacious and the effect of treatment is dependent on the time the patient has been treated, this model needs to be adopted so that time and baseline age are factors rather than the actual age of the child (which is the sum of the baseline age and the time in study). The results from the NH study suggest that the decline in DQ is not linear across all ages, and an interaction term time*baseline age is planned to be included in the model to allow for such a non-linearity. Also an unstructured variance-covariance matrix is planned to be used to account for the dependency between observations within the same patient. As a further justification for the model, it should be noted that time, baseline age and time*baseline age are all significant factors in the NH study.

To represent the treatment effect without constraint to a specific relation across visits, a term treatment*time is planned to be included in the model. This effect is used to judge the treatment effect. The null hypothesis of no treatment effect will be tested with a significance test for the treatment*time-effect in the second step of the multiple testing procedure (see Section 14.3). The difference between 104 weeks and baseline will be used as the contrast of interest for the test.

All missing values will be imputed. For a patients who stops treatment, let *tp* denote the turning point, i.e., the time point when the patient stops treatment. The patient's assessments as well as the time points for prediction are classified as being on either side of the turning point or at the turning point: prior to *tp*, at *tp*, or after *tp*. For patients who does not stop treatment, let $tp = \infty$, i.e., every time point is considered as prior to the *tp*. Let v denote a missing value to be imputed for a patient. The imputations will be based on two categories of predictions.
1). If the patient has two or more observations on the same side of *tp* as v (including a possible observation at *tp*), then v is considered as *exclusively patient-predictable (p_1)*, i.e., it can be predicted by simple linear regression based only on the patient's data (intra-or extrapolation). On the other hand, if only one or zero observations exists on the same side of *tp* (from the particular patient) as v (including a possible observation at *tp*), then it is considered as *non-predictable exclusively* from the patient's data.
2) The other type of prediction (p_2), will be calculated by the use of a population derived slope, estimated based on the repeated measures model in Section 14.1 (without imputations), and a patient observation. There will be two cases: If the patient has one or more observations on the same side of *tp* as v (including a possible observation at *tp*), then the nearest observation will be used together with a slope, which is based on the population (treated or untreated) of the same

status as the side of *tp* where the prediction time point is situated, as well as the interval relative the visits which contain the prediction time point. The other case is when there is no observation on the same side of *tp* as v (and no observation at *tp*). In this case, a prediction at *tp* will first be calculated, based on the same method as described above (under 2) but on the other side of *tp*, i.e., the nearest observation relative *tp* on the other side of *tp* will be used together with a slope based on the population (treated or untreated) of the same status as the side of *tp* where this nearest observation time point is situated, as well as the interval relative the visits which contain the *tp*. This prediction at the *tp* is in then used to predict v as described under the first case (under 2).

For each missing value v, the following algorithm is used to impute v by *i.*

If p_1 does not exists, then p_2 is used, i.e., i = p_2.

Otherwise, a weighted average will be used for the imputation: $i = w_1 p_1 + w_2 p_2$, where $w_1 = exp(-\lambda ln(2)/3)$ and $w_2 = 1 - exp(-\lambda ln(2)/3)$ and λ is the distance in months between the time point for prediction and the nearest available non-missing patient assessment time point. The exponential function is motivated by an exponential decrease of relevance of individual patient data relative to population data with increasing time, and the coefficient *ln(2)/3* is chosen such that population data reach the same weight as individual patient data after 75 % of the distance between two visits, i.e., three months. When individual patient data is available near a missed visit, the patient data weight is near 100 %, and when the distance is near one year on the other hand, the individual patient data weight has decreased to nearly 6 %.

14.1.1 Sensitivity analyses of the DQ endpoint

In order to assess the sensitivity of the model, the DQ model described in Section 14.1 will

- Only use non-missing values, i.e., no imputations will be performed
- Be extended with a triple interaction factor time*treatment*baseline age
- Use DQ = 0 for missing DQ values

An addition sensitivity analysis will be a tipping point analysis: All combinations of DQ values (n_{UN} , n_T) where n_{UN}, $n_T \in \{0,1,2,\ldots,100\}$ are substituted for missings in the untreated arm and the treated arm respectively. P-values will be calculated for each combination.

14.1.2 Pincipal strata analysis

A principal strata analysis will be performed where only patients without protocol violations will be analyzed according to Section 14.1.1. In particular patients who did not comply with the dosing scheme but continued the study will excluded from the analysis.

14.2 The HS endpoint

One of the key secondary objectives is to evaluate the effect of SOBI003 on HS levels in CSF.

To evaluate the effect of SOBI003 on HS levels in CSF linear models are used to model the change from baseline in HS levels as dependent variable, for both logged HS levels and untransformed HS levels, and baseline level and age as continuous covariates, and accumulated dose (including the FIH study) as covariate and sex as factor. The analyses are conducted for each assessment time point: baseline, week 52 and 104.

In addition, linear analyses with HS levels, both logged HS levels and untransformed HS levels, in CSF, serum and urine, as dependent variable, and baseline level and age as continuous covariates, and accumulated dose as covariate and sex as factor will be performed across all assessments (including the FIH study), with assessment as a repeated factor.

For the purpose of the multiple testing procedure (see Section 14.3), the linear repeated measures analysis with untransformed HS in CSF levels will be used, identical to the DQ analysis with the exception of the treatment factors.

The time effect is used to judge the HS level change in CSF. The null hypothesis of no change will be tested with a significance test for the time-effect in the first step of the multiple testing procedure (see Section 14.3). The difference between 104 weeks and baseline will be used as the contrast of interest for the test.

The evaluation of the effect of SOBI003 on HS levels in serum and urine will utilize the same models as mentioned above for HS levels in CSF.

The HS levels including change from baseline will also be summarized using descriptive statistics: n, mean, standard deviation (SD), standard error of the mean (SE), median, minimum and maximum. The HS change from baseline will also be presented as percentages. In addition, HS levels, change from baseline in HS levels and change from baseline in HS levels expressed as percentages will be presented in figures: by subject as well as mean graphs.

14.2.1 Sensitivity analyses of the HS endpoint

In order to assess the sensitivity of the model, the HS model described in 14.2 will

- Only use non-missing values, i.e., no imputations will be performed

An addition sensitivity analysis will be a tipping point analysis: All observed HS values in the study are substituted for missings.

14.3 Multiple comparison procedure

A hierarchical approach will be used for the key secondary co-endpoints.

The primary analysis will be based on a 2-sided test using a significance level of 0.05. A stepwise sequential testing procedure will be used to ensure a multiple level of significance of 0.05.

- 1st step: The null hypothesis of no difference between the treatments with respect to the reduction from baseline HS in CSF at 24 months will be tested using a significance level of 0.05. If the null hypothesis is rejected, then it will be concluded that SOBI003 reduces HS. Furthermore, the 2nd step of the sequential testing procedure will be performed.
- 2nd step: The null hypothesis of no difference between the treatments with respect to the change in DQ score at 24 months will be tested using a significance level of 0.05. If the null hypothesis is rejected, it will be concluded that Sobi 003 improves cognitive function as compared to untreated patients.

This multiple comparison procedure controls that the multiple level of significance is no more than 5 %.

Formally, the second confirmatory step will only be valid after the 1st step has been successfully met.

14.4 Adaptive Behavior

The adaptive behavior age-equivalent score is determined as the mean of the subdomain age-equivalent scores including the communication, socialization, and daily living skills (thus excluding the motor skills domain).

Adaptive behavior will be analyzed with the same repeated measures model as DQ and described in Section 14.1, but adapted for one group (i.e., no treatment*visit interaction term).

14.5 The MRI endpoints

As listed in Section 3.5.4, the MRI endpoints refers to changes in gray matter volume, compound ventricular volume, FA and MD in corpus callosum and cerebral white matter, SWI in cerebral white matter and basal ganglia.
To assess the PD effect of SOBI003 on MRI, linear models are used to model the change from baseline in MRI as dependent variable, for both logged MRI and untransformed MRI, and baseline level and age as continuous covariates, and accumulated dose (including the FIH study) as covariate and sex as factor. The analyses are conducted for each assessment time point: week 52 and 104.
In addition, linear analyses with MRI levels, both logged MRI levels and untransformed MRI levels, as dependent variable, and baseline level and age as continuous covariates, and accumulated dose as covariate and sex as factor will be performed across all assessments (including the FIH study), with assessment as a repeated factor.

14.6 Exploratory endpoints

Continuous variables are summarized using the number of patients, the mean, the standard deviation, the median, the minimum value, and the maximum value. Categorical variables are summarized using frequency counts and percentages.

14.7 Other efficacy endpoints

Endpoints relating to the neurocognition, liver and spleen volume, QoL, language, motor function and sleep pattern are summarized using descriptive statistics.

14.8 Subgroup Analyses

In order to examine the uniformity of the treatment effect across subgroups, additional exploratory statistical analyses will be performed on the DQ. The subgroups will be categorized according to:

- Baseline DQ
- Baseline AEq
- Sex (male, female)

The corresponding factors and treatment interactions will be included in the model, one at a time due to the small sample size.

No formal hypothesis tests will be performed.

14.9 Analysis of the subpopulation to be included in the future primary analysis population

The primary analysis will be repeated for the subgroup of subjects that will be included in the Phase III study (see Section 17), i.e., the subpopulation of the FAS patients who are up to 30 months of age and in the FIH study was initially assigned a dose at least as high as the dose selected for the phase III study, and the NH study patients who are up to 30 months of age. In the case the model does not converge due to few number of patients, only descriptive statistics will be presented.

15 Safety analyses

15.1 Drug exposure

The scope of the drug exposure will include the preceding FIH study as well as the present study. The presentations will be based both on the extension study separately, and pooled with the preceding study.
For each patient, graphs will be produced displaying dose by time and accumulated substance over time.
Duration of exposure will be calculated for each patient. It is defined as the total number of days a patient is exposed to study drug, and will be calculated using the following formula:
Time on treatment (days) = [(Treatment end date) – (Treatment start date) + 1].
The duration (days) and amount (mg) of exposure as well as the area under exposure curve will be summarized in a table using summary statistics, including the number of patients, the mean, standard deviation, median, minimum and maximum.

15.2 Adverse events

An adverse event (AE) is defined as any untoward medical occurrence, including a clinically significant laboratory finding, symptom, or disease in a patient enrolled into this study regardless of its causal relationship to study drug.

A treatment-emergent AE (TEAE) is defined as any event not present before exposure to study drug or any event already present that worsens in either severity or frequency after start of treatment.

Adverse events will be coded using the latest version of Medical Dictionary for Regulatory Activities (MedDRA).

An overview table will present the following information.

- Number and percentage of patients reporting TEAEs
- Number and percentage of patients reporting serious TEAEs
- Number and percentage of patients reporting TEAEs leading to death
- Number and percentage of patients reporting TEAEs leading to study discontinuation
- Number and percentage of patients reporting TEAEs related to study drug

The number of events will also be displayed in this summary table.

15.2.1 Adverse Events

Treatment emergent AEs will be summarized by system organ class, preferred term and ADA-status (positive, negative, boosted, unknown; each AE is classified according to the patients

ADA-status at the time of the AE occurrence;) displaying the number and percentage of patients with at least one AE recorded at each level of summarization. The total number of events and the number and percentage of patients with at least one TEAE will also be displayed. The summarization will be performed for the total safety analysis set and for the pivotal sub group. In addition, figures will be presented for each patient, displaying an overlay of drug exposure over time and adverse events marked out on the exposure curve. There will be two variants of this figure: one with dose by time, and the other with accumulated substance by time.

15.2.2 Serious adverse events

A serious adverse event (SAE) is defined as any event that:

- Results in death.
- Is life-threatening (i.e., at immediate risk of death).
- Requires in-patient hospitalization or prolongation of existing hospitalization.
- Results in persistent or significant disability/incapacity.
- Is a congenital anomaly/birth defect (i.e., in an offspring to the study patient).

SAEs will be summarized by system organ class and preferred term displaying the number and percentage of patients with at least one SAE recorded at each level of summarization. The total number of events and number and percentage of patients with at least one treatment emergent SAE will also be displayed.

15.2.3 Adverse Events leading to study discontinuation

Treatment emergent AEs leading to study discontinuation will be summarized by system organ class and preferred term displaying the number and percentage of patients with at least one AE recorded at each level of summarization. The total number of events leading to study discontinuation and the number and percentage of patients with at least one treatment emergent AE leading to study discontinuation will also be displayed.

15.2.4 Adverse Events by Severity

Treatment emergent AEs will be summarized by system organ class, preferred term and severity (mild, moderate, severe). The summarization will be performed for the total safety analysis set and for the pivotal sub group. The table will display the number and percentage of patients with at least one TEAE recorded at each level of summarization within the preferred term. The total number of events along with the number and percentage of patients with at least one treatment emergent AE will also be displayed by severity. In the event of missing severity information, the severity will be displayed as ‘Missing’.

The same presentation will be displayed for treatment emergent SAEs.

15.2.5 Adverse Events by Relationship to Study Drug

Treatment emergent AEs will be summarized by system organ class, preferred term and relationship to study drug (related, unrelated). The summarization will be performed for the total safety analysis set and for the pivotal sub group. The table will display the number and percentage of patients with at least one drug-related TEAE recorded at each level of summarization. The total number of events along with the number and percentage of patients with at least one drug-related treatment emergent AE will also be displayed.

Treatment emergent adverse events as judged by investigator as possible, probable or definite will be classed as related. Those recorded as not related or unlikely will be classed as unrelated. In the event of missing relationship information, the relationship will be imputed as related.

SAEs will be summarized in the same way.

15.2.6 Deaths

Information on the number and percentage of patients reporting TEAEs leading to death will be included in the overview table described in Section 15.2. Information on each death will be summarized in individual patient narratives.

15.3 Laboratory data

Laboratory test results will be compared to reference ranges, and values outside of the applicable reference range will be flagged as high (H) or low (L). In the data listing clinically significant values will also be indicated.

Laboratory data will be summarized by parameter and visit with descriptive statistics for the total safety analysis set and the pivotal sub group. This summary will be repeated within baseline ADA positive and baseline ADA negative subgroups.

Laboratory test results categorized as in or outside of the applicable range will be summarized using shift tables and shift plots i.e. shift from Baseline.

The concentration/value for each laboratory parameter will be plotted against time using a spaghetti plot for each patient.

Laboratory data will be listed in full.

15.4 Vital signs

Vital signs (blood pressure, heart rate, body temperature, respiratory rate and pulse oximetry)) are measured at the time points specified in the study schedule of events.

Observed and changes from baseline in vital signs will be summarized by visit with descriptive statistics for the total safety analysis set and the pivotal sub group.

The vital signs will be plotted using spaghetti plots as described for the laboratory data except without boundaries for normal ranges.

Vital signs results will be listed in full.

16 Pharmacokinetics and Immunogenicity

Blood sampling for PK, PD and immunogenicity is undertaken at the time-points specified in the study schedule of assessments.

16.1 Pharmacokinetic Parameters

The PK analysis will be based on the PK analysis set.

Serum concentrations will be listed and summarised by time point. Individual serum concentration profiles will be plotted both on the original scale and on the log scale. Mean (±SD) and median serum concentration profiles will be plotted for each dose both on the original scale and on the log scale. The individual serum concentration data, and the actual time for IMP administration and blood sampling will be used in the derivation of the PK parameters.

PK calculations, based on serum concentrations, will be performed in Phoenix WinNonlin by means of NCA. The following multiple-dose PK parameters will be determined for SOBI003 at Weeks 39, 52, 78 and 104, and during dose-adjustments:

• The time of the end of the infusion of SOBI003, tEnd of inf

• The observed serum concentration at the end of infusion of SOBI003, CEnd of inf

• The observed serum concentration immediately before the start of infusion of SOBI003, CPre-dose

• The area under the plasma concentration-time curve from time 0 to last sample, AUC168h

- Clearance, CL

AUC168h will be calculated according to the linear log trapezoidal method. Other PK parameters may be calculated as applicable.

In addition to evaluating SOBI003 PK based on SOBI003-002 data only, PK of SOBI003 will also be evaluated based on study SOBI003-002 together with PK data obtained from study SOBI003-001, since administration of SOBI003 have been started in study SOBI003-001 for all patients.

Pre-dose SOBI003 concentrations in study SOBI003-002 will be used to explore if steady-state is maintained or if SOBI003 serum concentrations decrease or increase following totally 2 years administration of SOBI003. The extent of accumulation of SOBI003 serum concentrations during 2 years treatment will be explored by calculating accumulation ratio (Rac) based on PK data from both study SOBI003-001 and study SOBI003-002. Rac will be calculated as AUC0-168h week n/AUC0-168h week 1, where n is 38, 52, 78 and 104.

The ratio of peripheral/central serum concentrations of SOBI003 may be explored in study SOBI003-002, in case SOBI003 serum concentrations have been determined in this study, depending on the outcome of study SOBI003-001.

The effect of eg. dose-level, age or body-weight on SOBI003 PK will be explored eg. by plotting PK parameters versus dose, age, body-weight.

The PK parameters will be listed for each patient and summarized for each IMP dose using descriptive statistics (the number of patients, arithmetic mean, standard deviation, standard error, coefficient of variation (CV) expressed as a percentage, geometric mean, geometric CV%, median, minimum, maximum). This summary will be repeated for PK parameters within baseline ADA positive and baseline ADA negative subgroups, within ADA positive at any postbaseline visit and ADA negative at all postbaseline visits.

16.2 Immunogenicity

The analysis of immunogenicity will in general be based on the immunogenicity analysis set. However, exceptions are the analyses of PK by ADA status at baseline, which will be based on the PK analysis set, and the AE analysis with ADA-categorized AEs in section 15.2.1.

For immunogenicity, baseline ADA (pre-existing ADA) from the previous study will be used.

The following definitions will apply as defined in the harmonised terminology for immunogenicity (Shankar et al., 2014, “Assessment and Reporting of the Clinical Immunogenicity of Therapeutic Proteins and Peptides—Harmonized Terminology and Tactical Recommendations”, The AAPS Journal, DoI:10.1208/s12248-014-9599-2):

- Evaluable patient: A patients with at least one sample taken after IMP administration that is appropriate for ADA testing (with reportable result).
- ADA-positive sample: When ADA is detected in a sample, the sample is considered positive.
- Baseline ADA: pre-existing antibodies detected at baseline.
- Treatment induced ADA: de novo development following administration of IMP (i.e., formation of ADA any time after the initial IMP administration in a patient without pre-existing ADA).
- Treatment boosted ADA: Baseline ADA boosted to higher level following IMP administration (i.e., at any time after the initial IMP administration the ADA titer is greater than the baseline titer by a factor of four).
- ADA-positive patient: A patient with at least one treatment-induced or treatment-boosted ADA-positive sample at any time during the treatment or follow-up observation period.
- ADA-negative patient: A patient without a treatment induced or treatment-boosted ADA-positive sample during the treatment or follow-up observation period.
- Baseline ADA Positive Patient: Patient with ADA at baseline. Such patient can have treatment boosted ADA after dosing.

Note that the requirement for inclusion in the immunogenicity set of at least one post-dose ADA assessment means that all patients in the immunogenicity set are “evaluable patients”.

The immunogenicity data will be listed for each patient and individual patient plots of titer measurement over time presented. Titer measurements will be summarized by visit and time point using descriptive statistics by IMP dose and overall. Immunogenicity will be evaluated in regards antibody prevalence (pre-existing antibodies), antibody incidence, titer and boosting following dosing, impact on pharmacokinetics responses.

A summary of ADA positive patients over time will be presented. For each time-point this will summarize the number of patients who were ADA positive (treatment induced or treatment boosted ADA). A summary of ADA change from baseline immune response will be presented, summarizing the overall ADA incidence (number of ADA positive patients) as well as the numbers of patients for whom this was treatment induced ADA and treatment boosted ADA, together with summaries of titer.

For the summary of treatment emergent AEs according to system organ class, preferred term and ADA-status, refer to section 15.2.1.

Figures for each patient will be produced with titer and AEs plotted against time with ADA status (positive, negative, boosted) will be marked as well.

The interval [0 TiterMax] that bounds the Titer values will be partitioned into up to four equal length intervals, where TiterMax is the maximal observed titer value. For each patient, each AE belonging to that patient will be classified according to which subinterval the nearest, previous titer belongs to when the AE occurred. For each titer-subinterval an AE summary for the SAF will be produced according to the analysis in 15.2.1 but without ADA-status. If one or more of the four subintervals are empty of corresponding AEs, then only three intervals will be used. If one or more subintervals of those three intervals are empty of corresponding AEs, only two subintervals will be used.

The same analysis will be performed for logged titers, where the interval [0 log(TiterMax)] that bounds the log(Titer) values will similarly be partitioned into up to four equal lengths intervals, where log(TiterMax) is the maximal observed titer value.

17 The primary analysis for the pivotal phase 3 study

In the Phase III study, patients up to 30 months are planned to be included. Therefore a subpopulation of the FAS patients who are up to 30 months of age and in the FIH study was initially assigned a dose at least as high as the dose selected for the phase III study will be included in the primary analysis population for demonstrating efficacy using combined data from the FIH study and the extension study, the Phase III trial and the corresponding patients in the NH study (control patients), consisting of the 5 patients with a rapid progressive MPS IIIA disease and a baseline age of 12 to 30 months. The primary analysis methods for DQ and HS and the estimand for the pivotal phase 3 study is planned to be identical to the corresponding analysis and estimand as described in the present SAP.

18 References

1 Shapiro EG, Nestrasil I, Delaney KA, Rudser K, Kovac V, Nair N, et al. A prospective natural history study of mucopolysaccharidosis type IIIA. J Pediatr, 2016; 170: 278-287 e271-274.