Official Title: A Phase II Multi-Center, Randomized, Double-Blind, 24 Week, Parallel Group, Placebo-Controlled Study to Investigate the Efficacy and Safety of Balovaptan (RO5285119) in Children and Adolescents Age 5-17 With Autism Spectrum Disorder (ASD)

NCT Number: NCT02901431

Document Date: SAP Version 2: 14-May-2019

Technical Document detailing the Statistical Analysis Plan of Study BP30153 (aV1ation)

Title: A PHASE II MULTI-CENTER, RANDOMIZED, DOUBLE-BLIND, 24-WEEK, PARALLEL GROUP, PLACEBO-CONTROLLED STUDY TO INVESTIGATE THE EFFICACY AND SAFETY OF BALOVAPTAN (RO5285119) IN CHILDREN AND ADOLESCENTS AGE 5-17 WITH AUTISM SPECTRUM DISORDER (ASD)

Protocol Number: BP30153 (aV1ation)

Study Drug: Balovaptan

[redacted]

14 May 2019
Date

[redacted]

21 May 2019
Date

[redacted]

5 June 2019
Date

Version: 2.0

Date: 14-MAY-2019

Table of Contents

1 BACKGROUND

This document describes the methods of summarizing and analyzing the data collected in Study BP30153. The main purpose of this document is to describe the data handling rules, derivation rules and statistical analysis methods for the final analysis of the double-blind, 24-week, placebo-controlled period that will be performed by Biostatistics and Statistical Programming and Analysis and that will be reported in the Clinical Study Report. An independent document will be prepared to describe the methods that will be used to summarize the open-label extension (OLE) data. Rules and methods here described will be applied, as appropriate, for any efficacy and safety interim analysis allowed by the study protocol.

Details on the PK and PK/PD analyses performed by the Pharmacometrics group within Clinical Pharmacology may be reported in a document separate from the Clinical Study Report.

This version 2.0 has been developed based on version 1.0, issued on 6-Aug-2018, on study protocol BP30153 version 6, issued on 19-Dec-2018, and on Study File Note "Dose Adjustment Following Review of PK Data by the IMC SOC" issued on 19-Jan-2019.

2 STUDY DESIGN

Study protocol BP30153 version 1-5 were conceived as "A PHASE II MULTI-CENTER, RANDOMIZED, DOUBLE-BLIND, 24-WEEK, **3-ARM, PARALLEL GROUP**, PLACEBO-CONTROLLED STUDY TO INVESTIGATE THE EFFICACY AND SAFETY OF *BALOVAPTAN (*RO5285119*)* IN CHILDREN AND ADOLESCENTS AGE 5-17 WITH AUTISM SPECTRUM DISORDER (ASD)", and version 6 as "A PHASE II MULTI-CENTER, RANDOMIZED, DOUBLE-BLIND, 24-WEEK, **PARALLEL GROUP**, PLACEBO-CONTROLLED STUDY TO INVESTIGATE THE EFFICACY AND SAFETY OF BALOVAPTAN (RO5285119) IN CHILDREN AND ADOLESCENTS AGE 5-17 WITH AUTISM SPECTRUM DISORDER (ASD)" with recruitment to the balovaptan 4mg equivalent treatment arm closed.

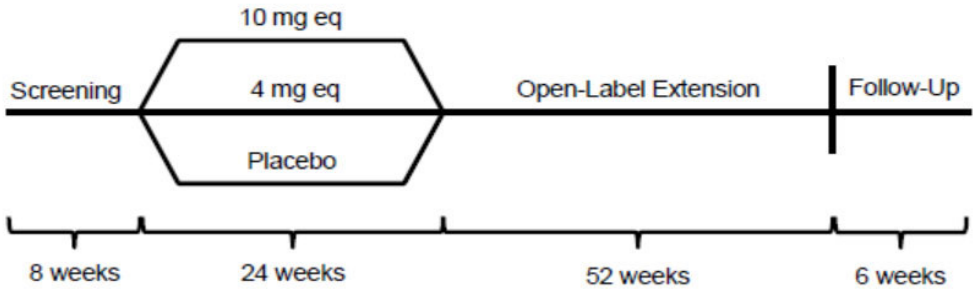
The overall study design was built by staggering enrollment by subject's age, starting first with adolescents (aged 13 to 17 years) and then with children (aged 5 to 12 years). Once the Internal Monitoring Committee (IMC) and the Scientific Oversight Committee (SOC) agreed on acceptable safety and tolerability in the first adolescent cohort (based on data from 8 weeks of treatment, First Study part) and determined the final doses, enrollment of adolescents resumed (for 24 weeks of treatment, Main Study part) and enrollment of a first cohort of younger subjects commenced. Similar process occurred for the first children cohort. See the Study Protocol for more details.

Approximately 340 children and adolescents aged 5 – 17 years with ASD are expected to be recruited to ensure a total of 160 evaluable subjects with placebo or 10 mg equivalent after 24 weeks of treatment.

Subjects who complete the double-blind 24-week treatment period are allowed to participate in an optional 52-week Open-Label Extension (OLE) period where they receive open-label balovaptan treatment (from Week 24 to Week 76).

For all subjects enrolled prior to protocol version 6, study design can be visualized as follows:

Figure 1. Study Design for subjects enrolled up to Version 5



Doses were initially administered as follows:

Table 1. Age-adjusted Starting Doses Based on PBPK Model

Age (years)	4 mg eq* Dose (mg)	10 mg eq* Dose (mg)
5 - 7	1.5	3
8 - 11	2	5
12 -14	3	7
15- 17	4	10

*e.q., dose predicted to achieve exposures to 4 mg/day or 10 mg/day, respectively, in adults

During the study, at planned IMC and SOC reviews of accrued data, the starting doses in Table 1 were shown to not achieve the target exposure in children aged between 5 and 14 years. Therefore, new doses were identified to produce exposures equivalent to the 4 mg and 10 mg doses used in adults (Study BP28420) as described by the following Table:

Table 2. Updated Age-adjusted Doses Based on PK Data

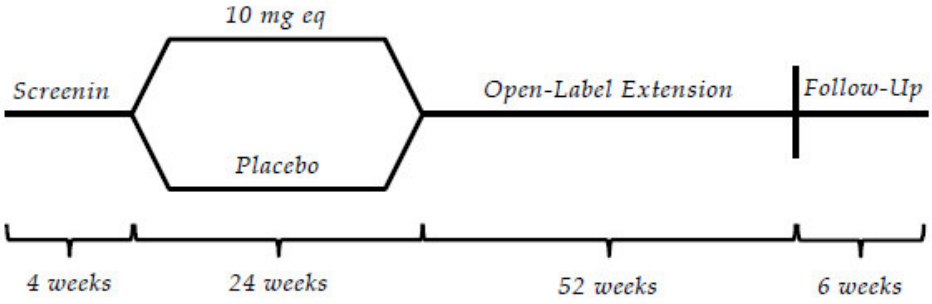
Age (years)	*4 mg eq Dose (mg)*	*10 mg eq Dose (mg)*
5–7 [a]	*3*	*7*
8–17	*4*	*10*

eq = age adjusted dose equivalent to adult dose; PK = pharmacokinetic.

[a] *Dose predicted to achieve exposures equivalent to 4 mg/day or 10 mg/day, respectively, in adults*

For subjects enrolled in accordance to protocol version 6, study design can be visualized as follows:

Figure 2. Study Design for subjects enrolled in accordance to Version 6



eq = equivalent.

For subjects randomized in accordance to protocol version 5 but after the 1st of October 2018 the dose is given, once the local IRB approved the plan described in the "Dose Adjustment Following Review of PK Data by the IMC SOC", at the age-adjusted dose level as described in Table 2.

For subjects still on-treatment in accordance to protocol version 5 after the 1st of October 2018 the dose is increased, once the local IRB approved the plan described in the "Dose Adjustment Following Review of PK Data by the IMC SOC", to the age-adjusted dose level as described in Table 2.

The following Table describes the disposition of subjects in relationship with study treatment in the Main study part, 24-weeks period.

Table 3. Study Treatments in the 24-weeks Main Part (Study Part 2)

Age (yrs)	Treatments randomized (1:1:1) according to Protocol Version 1-5			Treatments randomized (1:1:1) according to Protocol Version 5 and since 1st Oct 2018 allowed to start as (*) or to increase to (→) the Adult dose after local IRB approval				Treatments randomized (1:1) according to Protocol Version 6	
	Plc	4 mg eq	10 mg eq		Plc	4 mg eq	10 mg eq	Plc	10 mg eq
5-7	Plc	1.5 mg	3 mg	Treatment started as:	Plc*	3 mg*	7 mg*	Plc	7 mg
				Treatment increased to:	Plc → Plc	1.5 mg → 3 mg	3 mg → 7 mg		
8-12	Plc	2 mg	5 mg	Treatment started as:	Plc*	4 mg*	10 mg*	Plc	10 mg
				Treatment increased to:	Plc → Plc	2 mg → 4 mg	5 mg → 10 mg		
13-14	Plc	3 mg	7 mg	Treatment started as:	Plc*	4 mg*	10 mg*	Plc	10 mg
				Treatment increased to:	Plc → Plc	3 mg → 4 mg	7 mg → 10 mg		
15-17	Plc	4 mg	10 mg		Plc*	4 mg	10 mg	Plc	10 mg

The primary study objective (balovaptan 10 mg equivalent compared to placebo) will be addressed by analyzing the dataset obtained by pooling data from subjects that took placebo or balovaptan 10 mg equivalent, as highlighted in orange in Table 3, and belonging to patterns c) or d), as highlighted in yellow in Table 4 (see Section 4. Statistical Methods for more details).

2.1 OBJECTIVES

2.1.1 Primary Objective

The primary objectives of this study is:

- To evaluate the efficacy of 24-week treatment with balovaptan (RO5285119) 10 mg equivalent compared to placebo as measured by the change from baseline on the Vineland™-II Adaptive Behavior Scales, second edition (Vineland™-II) Two Domain Composite (2DC) (average of Communication and Socialization domains)

2.1.2 Secondary Objectives

The secondary objective of this study are as follows:

- To evaluate the efficacy of treatment with balovaptan 10 mg equivalent vs. placebo on:
  - Change from baseline on the Vineland™-II Composite standard score after 12 weeks and 24 weeks of treatment
  - Change from baseline in the Vineland™-II Communication, Socialization and Daily Living Skills domain standard scores after 12 weeks and 24 weeks of treatment
  - Proportion of subjects with ≥6-point improvement in the Vineland™-II 2DC score (clinically meaningful response)
  - Change from baseline in severity of clinical impressions as measured by CGI-S (Clinical Global Impressions-Severity) and OACIS-S (Ohio Autism Clinical Impressions Scale-Severity) after 12 weeks and 24 weeks of treatment
  - Improvements in clinical impressions as measured by CGI-I (Clinical Global Impressions-Improvement) and OACIS-I (Ohio Autism Clinical Impressions Scale-Improvement) after 12 weeks and 24 weeks of Treatment
  - Change from baseline in patient- or parent-reported Pediatric Quality of Life (PedsQL) v4.0 Generic Core Scale after 12 weeks and 24 weeks of treatment
  - Change from baseline in the Vineland™-II Composite standard score in adolescents and children independently after 12 weeks and 24 weeks of treatment
  - Change from baseline on the Vineland™-II 2DC score after 12 weeks of treatment
- To evaluate safety and tolerability of 24 and up to 76 weeks of treatment with balovaptan
- To evaluate the pharmacokinetics and exposure-response relationships of balovaptan and its metabolites, if appropriate

2.2 OUTCOME MEASURES

2.2.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline at Week 24 on the Vineland™-II Adaptive Behavior Scales 2-Domain Composite (2DC) Score, defined as the mean of the Communication domain standard score and the Socialization domain standard score.

2.2.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are as identified in the list of secondary objectives (see Section 2.1.2)

2.2.3 Exploratory Efficacy Endpoints

Exploratory efficacy endpoints are as follows:

- Change from baseline in behaviors as measured by Aberrant Behavior Checklist (ABC) Lethargy/Social withdrawal subscale after 12 weeks and 24 weeks of treatment
- Change from baseline in repetitive behaviors as measured by Repetitive Behavior Scale-Revised (RBS-R) after 12 weeks and 24 weeks of treatment
- Change from baseline to Week 76 (52 weeks of open-label treatment) as measured by Vineland™-II 2DC score
- Change from Week 24 to Week 76 (52 weeks of open-label treatment) as measured by Vineland™-II 2DC score
- Proportion of subjects with ≥4-point improvement in Vineland™-II 2DC score
- Proportion of subjects with ≥8-point improvement in Vineland™-II 2DC score
- Data (social symptoms, communication, and behavior) from Exit Interviews
- Change from baseline in behaviors as measured by ABC Irritability, Stereotypic Behavior, Hyperactivity/Noncompliance, and Inappropriate Speech subscales after 12 weeks and 24 weeks of treatment
- Change from baseline in patient- or parent-reported PedsQL Cognitive Functioning Scale and parent-reported PedsQL Family Impact Scale after 12 weeks and 24 weeks of treatment
- Change from baseline in caregiver-reported Global Impression (CaGI) severity on communication skills, social skills, daily living skills and overall autism symptoms after 12 weeks and 24 weeks of treatment
- Improvements in caregiver-reported Global Impression (CaGI) impression on communication skills, social skills, daily living skills and overall autism symptoms after 12 weeks and 24 weeks of treatment
- Palatability assessments (taste and acceptability)

2.2.4 Pharmacokinetic Endpoints

The pharmacokinetic endpoints are as follows:

- Apparent clearance (CL) and volume of distribution (VD)
- Exposure at steady-state (AUC0-24,ss)
- Other individual exposure estimates (e.g., Cmax, Tmax, AUC for a specific time interval) may also be derived as appropriate

2.2.5 Safety Endpoints

Safety will be assessed through:

- Occurrence, nature and intensity of adverse events, serious adverse events and non-serious adverse events of special interest, as determined using the Adverse Event Severity Grading Scale
- Physical and neurologic examinations, vital signs, hematology, blood chemistry and urinalyses, ECGs and C-SSRS

2.3 DETERMINATION OF SAMPLE SIZE

Prior to protocol version 6, a sample size of 240 subjects with ASD (80 per treatment arm) providing evaluable data at Week 24 ensured the study 80% power to detect as statistically significant, at 1-sided 5% significance level, a difference between each active dose and placebo with an effect size of at least 0.4. For the Vineland™-II 2DC score, assuming a standard deviation of about 12.5 points, then the effect size of 0.4 corresponds to 5 points. No adjustment for multiple doses was performed. Considering a withdrawal rate of around 15-20%, it was planned to recruit approximately 300 subjects overall.

In accordance to protocol version 6, 80 subjects per treatment arm (balovaptan 10 mg equivalent and placebo) are required, for a total sample size of approximately 160 subjects with ASD with evaluable data at Week 24. To achieve this number of evaluable subjects, the overall sample size in the study since it started is expected to increase to approximately 340 subjects.

2.4 ANALYSIS TIMING

The primary analysis will be conducted when the double-blind 24-week treatment period ends. Database lock will occur once all subjects have either completed the 24-week assessment or withdrawn from the study early, and all data required for analysis have been cleaned and verified.

An efficacy and safety interim analysis is planned once approximately 80 subjects taking either balovaptan 10 mg equivalent or placebo (i.e., approximately 40 subjects per treatment arm) have completed their 12-week visit without dose interruptions or adjustments to allow internal decisions for the next steps of the development plan.

Given the hypothesis generating nature of this study, the Sponsor may conduct up to two interim analyses beyond what is specified elsewhere in this protocol. The decision to conduct such an interim analysis and its timing will be documented in the Sponsor's study master file prior to the conduct of the interim analysis. The clinical study report will also document that such an interim analysis occurred. Interim analyses will be performed and interpreted by the IMC and SOC members (as appropriate), who will have full access to unblinded data.

3 STUDY CONDUCT

3.1 RANDOMIZATION CONSIDERATIONS

After all subjects have been randomized, the data entered into the IxRS system will be reviewed for consistency with the data entered into the Case Report Form (CRF). In

particular, the randomization dates and stratification factors (sex and age) will be checked and a summary of the discrepancies between the IxRS and CRF will be reviewed.

3.2 TREATMENT GROUPING

As the starting doses in children aged between 5 and 14 years were shown to not achieve the target adult exposure, the treatment grouping originally expected - Placebo, balovaptan 4 mg equivalent and balovaptan 10 mg equivalent - is no longer tenable.

Given the numerous different dose regimens occurred in the study, as described in Table 3, the individual treatment level will be ultimately identified based on individual subject's PK exposure, estimated as the average plasma concentration (C_{ave}) since treatment start.

C_{ave} was not specifically listed in the Pharmacokinetic endpoints section of the protocol, however, in the situation of possible dose adaptations within an individual during the course of the study, C_{ave} is the PK exposure measure that reflects most realistically the individual exposure. Moreover, C_{ave} also takes into account potential dose interruptions during treatment. For subjects without dose changes or dose interruptions, Cave is approximately equal to $AUC_{0\text{-}24,ss}$ divided by 24 hrs. Individual C_{ave} will be estimated by the pharmacometrics group in Clinical Pharmacology for all subjects with at least one adequately documented PK measurement at steady-state, using a population PK modeling approach. C_{ave} at week 12 will be used as a classification factor to split the subjects receiving balovaptan into three equally sized exposure sub-groups: Low Tertile, Medium Tertile and High Tertile.

If insufficient exposure data are available for a subject and C_{ave} cannot be derived (this can happen, for example, to withdrawals or in case of issues with PK sample analyses), then his/her data will contribute to the same treatment group of the majority of his/her coetaneous taking the same regimen as per randomisation. For example, if a 6 years-old initially randomized to 1.5mg switched to 3mg and no PK data are available for him/her, then his/her data will be attributed to the Low Tertile if most of the 5-7 years-old subjects taking the same regimen are allocated, based on their available exposure, to the Low Tertile. In case no other coetaneous providing exposure data went through the same regimen pattern (as in the example, increasing dose from 1.5mg to 3mg) then data will be attributed to the same exposure sub-group of the majority of his/her coetaneous taking the same final dose (in the example, 3mg).

This treatment grouping will be used for descriptive (ie, not inferential) statistical reporting purposes of efficacy and safety data.

For pharmacometric exploratory graphical analyses, missing PK information will not be inputed. In case of missing PK information for subjects assigned to a starting dose other than Placebo, the subject with missing PK information will be excluded from exposure-efficacy analyses.

3.3 TIME-WINDOWS

Efficacy and safety analyses will be performed according to the data collected at the originally scheduled visit.

In case of deviations from the visit date scheduled as per protocol the following boundaries will be implemented to identify time-windows:

Site Visit	Time-window (days)	Range (days)
Baseline	Before or soon after Start Dosing	Up to the 1st day of dosing
Week 6	43 ± 21	23 – 64
Week 12	85 ± 21	65 – 106
Week 18	127 ± 21	107 – 148
Week 24	169 ± 21	149 – at most 21 days after last double-blind Dose and prior to OLE Screening

For Screening, the scheduled visit will be used regardless of the actual assessment days relative to the first day of study drug.

Baseline Definition

Unless otherwise stated, baseline is defined as the last non-missing value recorded before the first study drug administration of each repeated study drug period (ie. 8 weeks for the First Part, and 24 weeks for the Main Part) or soon after the start of the dosing (see Table above). The last observation can be an unscheduled / repeated measurement.

Baseline for Electrocardiograms (ECGs) is derived as the mean of the triplicate measurements of pre-dose assessments.

Unscheduled Measurements

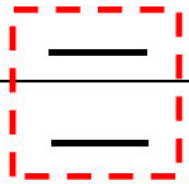
Unscheduled measurements will be included in the listings. Except for unscheduled measurements used for baseline, unscheduled measurements will be excluded from the descriptive statistics and statistical analysis.

4 STATISTICAL METHODS

4.1 ANALYSIS DATASETS

According to the protocol, the disposition of subjects in relationship with each Study Part may be as follows:

Table 4. Patterns of subjects' disposition

Pattern	**Description**: Subjects …	**First Part** (Study part 1): up to 8 weeks of Treatment	**Main Part** (Study part 2): 24 weeks of Treatment
a)	stopped treatment prior to determination of final doses by IMC/SOC		
b)	stopped treatment prior to determination of final doses by the IMC/SOC and took the opportunity to restart in the Main Part		
c)	allowed by the IMC/SOC to continue and complete the 24 weeks of treatment without dose interruption/adjustment		
d)	enrolled in the Main Part		

The primary **Efficacy** treatment comparison (balovaptan 10 mg equivalent dose vs. placebo) will be performed on the dataset obtained by pooling data from patients that were expected to take treatment for 24 weeks, as described by patterns c) and d) [highlighted in yellow].

Safety analysis will be performed separately on two datasets, according to the duration of the seamless treatment period that subjects have been exposed to. Therefore, two safety analysis datasets are identified:

- **First Part (8 weeks)** - Data from subjects that took treatment up to a maximum of 8 weeks, as described in pattern a) and 1st part of b) [framed with dotted red line]
- **Main Part (24 weeks)** - Data from subjects that took treatment up to a maximum of 24 weeks, as described in pattern c), d), and 2nd part of b) [framed with dotted blue line]

4.2 ANALYSIS POPULATIONS

Given the impact of the early findings from the starting doses on the actual doses regimens implemented in the trial, on the final treatment classification required for summary purposes and on the identification of the appropriate dataset to estimate the treatment comparison, the following populations are considered in order to provide an exhaustive reporting of this study data.

4.2.1 Safety Population

- *Safety Global*

The “Safety Global” population consists of all subjects who gave informed consent and received at least one dose of study medication, whether prematurely withdrawn from the study or not.

Data will be summarized according to treatment classified as derived from PK exposure: Placebo, Low Tertile, Medium Tertile, High Tertile and All Treated with Active.

- *Safety Inferential*

The “Safety Inferential” population consists of the subset of the “Safety Global” population that is identified to address the primary efficacy study objective.

Data will be summarized according to treatment classified as: Placebo, 10mg equivalent.

4.2.2 Efficacy Population

- *Efficacy ITT Global*

The “Efficacy Global” population consists of all subjects who gave informed consent and received at least one dose of study medication, whether prematurely withdrawn from the study or not.

Data will be summarized according to treatment classified as derived from PK exposure: Placebo, Low Tertile, Medium Tertile, High Tertile.

- *Efficacy ITT Inferential*

The “Efficacy ITT Inferential” population is the primary population of interest. It consistsof the subset of the “Efficacy Global” population obtained by pooling data from subjects taking balovaptan 10 mg equivalent or the concurrently randomized placebo in the corresponding age band and in the same randomization stage, as highlighted in orange in Table 3, and belonging to patterns c) or d), as highlighted in yellow in Table 4. Subjects with dose adjustments or interruptions, or on a different dose than the final dose for their age group, are excluded and do not contribute to this analysis population.

Data will be summarized according to treatment classified as: Placebo, 10mg equivalent.

The following Table shows the combination of subjects' patterns (described in Table 4) to identify the Safety and Efficacy analyses population/dataset:

Table 5. Patterns of subjects' disposition identifying Safety and Efficacy Analysis Population/Dataset

Analysis Population:	Analysis dataset:	
	First Part (8 weeks)	**Main Part (24 weeks)**
Safety Global	a + 1st part of b	2nd part of b + c + d
Safety Inferential	-	c + d
Efficacy ITT Global	-	2nd part of b + c + d
Efficacy ITT Inferential	-	c + d

4.2.3 PK/PD Population

The PK/PD population will be used for exploratory PK/PD analyses. The PK/PD population consists of a subset of the Efficacy Global population and includes all subjects with at least one post-baseline efficacy measurement and an exposure estimate. For subjects assigned to Placebo, exposure will be assumed to be 0. In case of missing PK information for subjects assigned to a starting dose other than Placebo, the subject with missing PK information will be excluded from exposure-efficacy analyses.

4.3 ANALYSIS OF STUDY CONDUCT

The number of subjects who enroll, discontinue or complete the study will be summarized for the Safety population/dataset by treatment group. Reasons for premature withdrawal from the study and premature withdrawal from treatment will be listed and summarized.

4.4 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Demographic and baseline characteristics (including age, sex, WASI-II IQ score, SRS-2 total t-score, Vineland™-II 2DC score) and clinically relevant disease characteristics (including background antipsychotics) will be summarized for the Safety population/dataset by treatment group using means, standard deviations, medians and ranges for continuous variables and proportions for categorical variables, as appropriate.

4.5 EFFICACY ANALYSES

Summary tables and graphs will be provided both for the "Efficacy ITT Global" and for the "Efficacy ITT Inferential" populations, while listings of individual data will be provided for the "Efficacy ITT Global" population only, separately by subjects' patterns (b, c, d).

Inferential statistical analysis will be performed on pooled data from subjects taking balovaptan 10 mg equivalent dose compared with those from the concurrently randomized placebo in the corresponding randomization stage, as identified by the "Efficacy ITT Inferential" population. Multiple endpoints will be analyzed however, due to the exploratory nature of the study, multiplicity will not be statistically adjusted for and the risk of false positive results will be taken into consideration during the interpretation of the results.

The primary efficacy endpoint, the change from baseline of Vineland™-II 2-DC standard score, as well as the secondary endpoints expressed in terms of change from baseline on a continuous scale (namely, Vineland™-II standard scores, PedsQL Scale factors, ABC factors and RBS-R factors) will be analyzed using Mixed Model Repeat Measurements (MMRM) on the overall population of adolescents and children. The MMRM model will include treatment, sex and visit (week) as main effects, individual age and baseline score as covariates, and treatment-by-visit and baseline-by-visit as interaction terms. Visit will be fitted as a repeated effect with an unstructured correlation structure across visits within each subject. All main effects and interactions will be retained in the final model regardless of their statistical significance.

The results of the analysis will be presented as point estimates, 90% confidence intervals and associated p-values for the adjusted mean differences between 10 mg eq balovaptan and placebo after 24 weeks of treatment as well as at intermediate visits.

The main analysis outlined above will be done using a mixed effect model, which can handle missing data without any need to recur to imputation or discarding of data.

The proportion of subjects with ≥6-point improvement in the Vineland™-II 2DC score (clinically meaningful response) will be analyzed using logistic regression. A Generalised Estimating Equations (GEE) model will be fitted including treatment, sex and visit (week) as main effects, age as covariate, baseline-by-visit and treatment-by-visit interactions, with subject fitted as a repeated effect. It will be based on the binomial distribution and thus the link function will be the 'logit'.

The odds ratio will be presented for descriptive purposes along with the corresponding 90% confidence interval and p-value for the comparisons between 10 mg eq balovaptan and placebo after 24 weeks of treatment as well as at intermediate visits.

Change from baseline in severity of clinical impressions as measured by CGI-S
The proportion of subjects with at least 1-point decrease (i.e. improvement) from baseline on CGI-S will be analyzed using logistic regression. A GEE model will be fitted including treatment, sex and visit (week) as main effects, age and baseline CGI-S as covariates, baseline-by-visit and treatment-by-visit interactions, with subject fitted as a repeated effect. It will be based on the binomial distribution and thus the link function will be the 'logit'.

The odds ratio will be presented for descriptive purposes along with the corresponding 90% confidence interval and p-value for the comparisons between 10 mg eq balovaptan and placebo after 24 weeks of treatment as well as at intermediate visits.

Improvements in clinical impressions, as measured by CGI-I scores
The proportion of subjects with a value of at least 2 (i.e., "Much improved" or better) on CGI-I will be analyzed using logistic regression applying the same model above described for CGI-S.

4.5.1 Subgroup Analyses

The results of the primary and secondary efficacy variables will be summarized within subgroups using descriptive statistics. The following categorization will be used to define the subgroups:

- Sex (male, female)
- Age (adolescents 13-17yrs, children 5-12yrs)

The statistical model for subgroup analyses of efficacy data will be performed using MMRM with subgroup, treatment-by-subgroup, subgroup-by-time and treatment-by-subgroup-by-time interaction terms included along the independent effects described above.

Estimates of treatment effect and associated 90% CIs will be presented in forest plots. Unadjusted p-values will also be presented for these analyses.

4.5.2 Interim Analysis

The populations assessed at the Interim Analysis will be the "Safety Inferential" and the "Efficacy ITT Inferential" with all subjects randomized prior to 21FEB2019, which is the date of the last subject expected to provide 12 week data for the 10mg-Placebo comparison by 16MAY19.

The efficacy interim analysis will focus on the outcome measures derived from Vineland™-II and, only descriptively, from PedsQL v4.0 Generic Core Scale, CGI-I and CGI-S.

Summary statistics will be provided for the overall population and for the adolescents/ children subgroups separately.

4.5.3 PK/PD Analyses

An exploratory graphical analysis of the relationship between pharmacokinetic exposure and efficacy will be performed by Clinical Pharmacology. For the exploratory graphical exposure-efficacy analyses, C_{ave} will be used either as a continuous covariate or as grouping variable. Graphical displays will include (but are not limited to) efficacy outcome measure vs. exposure by visit, and efficacy outcome measure vs. time, with trend lines split by exposure category (e.g. placebo, Low, Medium and High Tertile of C_{ave}). The effect of subpopulations will also be investigated.

If warranted after review of initial graphical displays, a model based approach may be considered to quantify the exposure-effect relationship for Vineland™-II 2-DC. A model based exposure-efficacy analysis of other efficacy outcome measures might be considered upon agreement with the project team.

Results from pharmacometric analyses may be reported in a document separate from the clinical study report.

4.6 SAFETY ANALYSES

Summary tables and graphs will be provided both for the "Safety Global" and for the "Safety Inferential" populations, while listings of individual data will be provided for the "Safety Global" population only, separately by subjects' patterns (a, b, c, d).

As appropriate, listings, summary tables and graphs by treatment group will be provided for safety and tolerability assessments, including:

- Incidence of adverse events (overall, by intensity, and by relationship to study medication).

- Serious adverse events and adverse events of special interest will be reported separately.
- [redacted]
- Vital signs (systolic blood pressure, diastolic blood pressure, pulse rate, and orthostatic effects)
- [redacted]
- Physical examination (height, weight, neurological examination)
- Safety laboratory values (including hematology, blood chemistry, coagulation, and urinalysis parameters)
- Incidence of marked laboratory abnormalities
- Clinical assessment of suicidality (C-SSRS)

Adverse Events

Verbatim descriptions of adverse events recorded on the eCRF by the Investigator during the study period will be mapped to the Medical Dictionary for Regulatory Activities (MedDRA) thesaurus terms. All adverse events will be tabulated by body system and preferred term for individual events within each body system. Adverse events will also be tabulated by severity and relationship to the study medication.

Serious adverse events will be summarized separately.

Clinical Laboratory

Subjects' listings and summary statistics by treatment group at each assessment time will be presented using the International System of Units (SI units; Système International d'Unités). Laboratory data not reported in SI units will be converted to SI units before processing.

Laboratory test values will be presented by individual listings with flagging of values outside the normal ranges Roche standard reference ranges, rather than the reference ranges of the Investigator, will be used for all parameters. For most parameters the measured laboratory test result will be assessed directly using the Roche standard reference range. A transformation will be performed on some laboratory tests that lack sufficiently common procedures and have a wide range of Investigator ranges, e.g., enzyme tests that include AST, ALT, and alkaline phosphatase and total bilirubin. If the standard reference ranges for these parameters have a lower limit of zero only the upper limits of the ranges will be used in transforming the data.

For all laboratory parameters included, there exists a Roche predefined standard reference range. Laboratory values falling outside this standard reference range will be labeled "H" for high or "L" for low in study subject listings of laboratory data.

In addition to the standard reference range, a marked reference range has been predefined by Roche for each laboratory parameter. The marked reference range is broader than the standard reference range. Values falling outside the marked reference range that also

represent a defined change from baseline will be considered marked laboratory abnormalities (i.e., potentially clinically relevant). If a baseline value is not available for a subject, the midpoint of the standard reference range will be used as the subject's baseline value for the purposes of determining marked laboratory abnormalities. Marked laboratory abnormalities will be labeled in the subject listings as "HH" for very high or "LL" for very low.

Vital Signs and ECG

Vital signs and ECG will be presented by individual listings and summary tables by treatment group and time of raw values and change from baseline.

Concomitant Medications

Concomitant medications will be presented in summary tables and listings.

5 DERIVED ENDPOINTS

- **Aberrant Behavior Checklist (ABC)**

  The ABC consists of 58 items subdivided in 5 subscales: Irritability, Lethargy and Social Withdrawal, Stereotypic Behavior, Hyperactivity/Non-Compliance and Inappropriate Speech.

  The total scores for each of the 5 subscales (but not the values of all 58 items) will be electronically transferred to Roche.

- **Pediatric Quality of Life Inventory (PedsQL ™ 4.0) Generic Core**

  The Pediatric Quality of Life Inventory PedsQL™4.0 Generic Core Scale assessment consists of 23 items encompassing 4 core scale domains:

  — Physical Functioning (8 items)

  — Emotional Functioning (5 items)

  — Social Functioning (5 items)

  — School Functioning (5 items)

  summarised into:

  — Total Score (23 items)

  To create the scores for the above described 4 Functioning Scales and the Total Score a mean is computed as the sum of the items divided by the number of items answered (this accounts for missing data). If more than 50% of the items in the scale are missing, the Scale Score is not computed [Viecili and Weiss, 2015].

  Additional Summary Scores will be derived as follows:

  — Psychosocial Health Summary Score, as the sum of the items divided by the number of items (15, in case of no missing values) answered in the Emotional, Social, and School Functioning Scales.

- Physical Health Summary Score, which is the same as the Physical Functioning Scale Score.
- PedsQL Total Score, computed as the sum of all the items divided by the number of items (23, in case of no missing values) answered on all the Scales.

If more than 50% of the items are missing, the Summary Score is not computed [Viecili and Weiss, 2015].

The above described Functioning Scales and Summary Scores will be derived by Roche Statistical Programmer.

- **Pediatric Quality of Life Cognitive Functioning Scale**

  The PedsQL Cognitive Functioning Scale consists of 6 items.

  To create the Cognitive Functioning Total Score a mean is computed as the sum of the items divided by the number of items answered (this accounts for missing data). If more than 50% of the items in the scale are missing, the Scale Score is not computed [Viecili and Weiss, 2015].

- **Pediatric Quality of Life Family Impact Scale**

  The Pediatric Quality of Life Inventory (PedsQL™) Family Impact Module version 2 consists of 36 items encompassing the following 6 scales

  - Physical Functioning (6 items)
  - Emotional Functioning (5 items)
  - Social Functioning (4 items)
  - Cognitive Functioning (5 items)
  - Communication (3 items)
  - Worry (5 items)

  and the 2 following scales measuring parent-reported family functioning

  - Daily Activities (3 items)
  - Family Relationships (5 items)

  The above described 6+2 scales are summarised by mean scores computed as the sum of the items scores divided by the number of items answered (this accounts for missing data). If more than 50% of the items in the scale are missing, the Scale Score is not computed [Varni et al., 2004].

  Additional Summary Scores will be derived as follows:

  - Total Score, by averaging across all 36 items
  - Parent HRQoL Summary Score, by averaging the 20 items in Physical, Emotional, Social and Cognitive Functioning
  - Family Summary Score, by averaging the 8 items in Daily Activities and Family Relationships

The above described Functioning Scale and Summary Scores will be derived by Roche Statistical Programmer.

- **Repetitive Behavior Scale - Revised (RBS-R)**

  The RBS-R is a 43-item informant-based questionnaire assessing the variety of restricted and repetitive behaviors observed in individuals with ASD. The scale is grouped into 6 subscales: Stereotyped, Self-injurious, Compulsive, Ritualistic, Sameness, and Restricted Behaviors.

  The subscales score will be electronically transferred to Roche.

- **Vineland™-II Adaptive Behaviour Scale (VABS)**

  Values of Vineland™-II Adaptive Behavior Composite Score (as Sum and Standard score), domains (as v-scale and standard score for each of the 3 domains, Communication, Daily Living Skills, Socialization), subdomains (as raw and v-scale for each of the subdomains) and age-equivalent values (years:months, for each subdomain) will be derived via computerized system and electronically transferred to Roche.

  The value for the primary endpoint, Vineland™-II Adaptive Behavior Scales 2-Domain Composite (2DC) Score, defined as the mean of the Communication domain standard score and the Socialization domain standard score, will be derived by Roche Statistical Programmer. If any of the two individual domain standard scores is missing the 2DC score is not computed.