STATISTICAL ANALYSIS PLAN

Final Version 3.0 dated 23-MAR-2023

A Phase 3, Double-blind, Randomized, Multicenter, Parallel Group, Placebo-controlled Sequential Dose Titration Study to Evaluate Efficacy, Safety and Pharmacokinetics of Mirabegron in Pediatric Subjects from 5 to <18 Years of Age with Overactive Bladder

ISN: 178-CL-204
EudraCT number: 2016-001767-37

Astellas Pharma Global Development Inc. (APGD)
1 Astellas Way
Northbrook, IL 60062, US

Table of Contents

List of In-text Tables

List of In-text Figures

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
AE	Adverse Event
ALP	Alkaline Phosphatase
ALT	Alanine Transaminase
ANCOVA	Analysis of Covariance
APGD	Astellas Pharma Global Development
ASCM	Analysis Set Classification Meeting
AST	Aspartate Transaminase
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CI	Confidence Intervals
CRF	Case Report Form
CS	Classification Specifications
CSR	Clinical Study Report
DBP	Diastolic Blood Pressure
DSMB	Data Safety Monitoring Board
EBC	Expected Bladder Capacity
ECG	Electrocardiogram
EOS	End of Study
EOT	End of Treatment
FAS	Full Analysis Set
GCMSL	Global Clinical Modeling & Simulation Lead
GD	Global Development
H	High
HTN	Hypertension
ICCS	International Children's Continence Society
ICF	Informed Consent Form
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IDAC	Independent Data Analysis Center
IP	Investigational Product
IRT	Interactive Response Technology
L	Low
LS	Least Squares
LLN	Lower Limit of Normal
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
ms	Millisecond
N	Normal
OAB	Overactive Bladder
PD	Pharmacodynamic

Abbreviations	Description of abbreviations
PD1-x	Protocol Deviation 1-x
PK	Pharmacokinetic
PKAS	Pharmacokinetics Analysis Set
PKDAP	Pharmacokinetic Data Analysis Plan
PPIUS	Patient Perception of Intensity of Urgency Scale
PPS	Per-Protocol Set
PT	Preferred Term
PVR	Post Void Residual
QTc	Corrected Q-T Interval
QTcF	QT interval corrected by Fridericia's formula
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SAS	Statistical Analysis Software
SBP	Systolic Blood Pressure
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TLF	Tables, Listings and Figures
UA	Urinalysis
ULN	Upper Limit of Normal
UTI	Urinary Tract Infection
WHO-DD	World Health Organization Drug Dictionary

List of Key Terms

Terms	Definition of terms
Baseline	Observed values/findings which are considered to be the starting point for comparison.
Discontinuation	The act of concluding participation in a trial by an enrolled subject, prior to completion of all protocol required elements. Note: subject discontinuation does not necessarily imply exclusion of subject data from analysis that was collected prior to discontinuation.
Enroll	To register or enter into a clinical trial, i.e., signing the informed consent form (ICF). Once a subject has been enrolled, the clinical trial protocol applies to the subject.
EoT result	End of treatment result is based on the last available result which is obtained earlier or on the same date as the treatment stop date.
Investigational period	Period of time where major interests of protocol objectives are observed, and where the study drug is given to a subject. This period continues until the last assessment after completing the last dose of the study drug.
Pediatric equivalent dose (PEDx)	Weight-range based doses predicted to achieve plasma concentrations equivalent to steady state exposures expected with "x" mg mirabegron administered once daily in adults.
Postinvestigational period	Period of time after the last assessment of the protocol. Follow-up observations for sustained adverse events and/or survival are done in this period.
Screening	A process of active consideration of potential subjects for a trial.
Screening period	Period of time before entering the investigational period, usually from the time the subject signed informed consent until just before the first dose of the study drug is given to a subject.
Screening failure	Screened subject who did not fulfill protocol inclusion and/or exclusion criteria, or decided not to participate anymore (withdrew consent) prior to first dose of study drug.
Source data	All information in original records or certified copies of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial. Source data are contained in source documents (original records, certified copies).
Source documents	Original documents, data, and records including source data.
Steady state	When the amount of drug intake is equilibrium with the rate of drug elimination. Note: for mirabegron steady state is considered to be reached after 10 days of daily dosing.
Study period	Period of time from the first site initiation date to the last site completing the study.
Subject	An individual in the population of interest who participates in a clinical trial as recipient of the investigational product.
Treatment emergent adverse event	An adverse event observed after starting administration of the study drug.
Trough sample	Pharmacokinetic sample taken just prior to the next dose of study medication.

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes procedures for executing the statistical analysis to fulfil the objectives of the study.

The final SAP will be approved prior to database hard lock and unblinding the subject treatment assignments. If needed, revisions to the approved SAP may be made prior to database hard lock. Revisions will be version controlled.

Changes from the planned analyses in the final SAP that impact the statistical analyses will be documented in the Clinical Study Report (CSR).

Prior to database hard lock, a final review of data and Tables Listings and Figures (TLFs) meeting will be held to allow a review of the clinical trial data and to verify the data that will be used for analysis set classification. If required, consequences for the statistical analysis will be discussed and documented. A meeting to determine analysis set classifications may also be held prior to database hard lock.

Pharmacokinetic (PK) analyses will be described in a separate data analysis plan (PKDAP) which will be finalized prior to the analysis commencement and will be reported separately.

2 STUDY OBJECTIVE(S) AND DESIGN

2.1 Study Objectives

The primary, secondary and exploratory objectives and endpoints for this study are listed in the table below. Primary and secondary objectives and endpoints apply to children only; exploratory objectives and endpoints also apply to adolescents.

Table 1 Study Objectives and Endpoints

Objectives	Endpoints
Primary	
● To evaluate the efficacy of mirabegron in children (5 to < 12 years of age) with Overactive Bladder (OAB)	● Change from baseline at the end of the 12-week treatment period: ○ Mean number of micturitions per 24 hours
Secondary	
● To evaluate the efficacy of mirabegron in children (5 to < 12 years of age) with OAB	● Change from baseline at the end of the 12-week treatment period: ○ Mean volume voided per 24 hours ○ Maximum volume voided ○ Mean number of daytime incontinence episodes per 24 hours ○ Mean number of daytime micturitions per 24 hours ● Number of dry (incontinence-free) days per 7 days at the end of the 12-week treatment period
Table continued on next page	

Objectives	**Endpoints**
● To evaluate the safety and tolerability of mirabegron in pediatric subjects with OAB	● Nature, frequency and severity of AEs ● Clinical laboratory tests (hematology, biochemistry and urinalysis) ● Vital signs (blood pressure and pulse) ● Routine 12-lead ECG ● PVR volume ● Acceptability and palatability questionnaire
● To evaluate the pharmacokinetics after multiple dose administration of mirabegron in pediatric subjects with OAB	● Steady-state C_{max}, AUC_{tau}, C_{trough}, T_{max}, CL/F and Vz/F.
Exploratory	
● To evaluate the efficacy of mirabegron in pediatric subjects (5 to < 18 years) with OAB	● Percentage of subjects with a reduction in daytime incontinence episodes (< 50% reduction [nonresponder], 50% [partial responder] and 100% [responder]) ● Improvement from baseline in worst incontinence grading ● Mean number of micturitions per 24 hours, adjusted for fluid intake: ○ Number of micturitions per 24 hours that would be required at week 12/EoT to void the same total volume as voided at baseline ○ Difference in mean number of micturitions per 24 hours adjusted for fluid intake compared to the expected mean number ● Mean volume voided per micturition as a percentage of expected bladder capacity ● Categorized mean volume voided per micturition as a percentage of expected bladder capacity (results and changes from baseline). ● Change from baseline at the end of the 12-week treatment period adjusted for fluid intake: ○ Mean number of micturitions per 24 hours ● Change from baseline at the end of the 12-week treatment period (adolescents only): ○ Mean number of micturitions per 24 hours ○ Mean number of daytime micturitions per 24 hours ○ Mean volume voided per 24 hours ○ Mean number of incontinence episodes per 24 hours ○ Maximum volume voided ● Number of dry (incontinence-free) days per 7 days at the end of the 12-week treatment period (adolescents only) ● Mean number of daytime grade 3 or 4 (PPIUS) urgency episodes per 24 hours (adolescents only)

AE: adverse event; ECG: electrocardiogram; OAB: overactive bladder; PPIUS: patient perception of intensity of urgency scale; PVR: post void residual.

2.1.1 Estimand

The estimand of most clinical importance for this study is defined by the following four attributes:

- Target population: all children who took at least 1 dose of the study drug, and in whom a non-missing measurement for micturition frequency at baseline and after administration of the study drug is available
- Outcome measurement: micturition frequency/day (24 hours)
- Intercurrent events:
  1. Discontinuation from treatment (if week 12/EoT not obtained)
  2. AE Urinary Tract Infection (UTI) (if occurred in week 12/EoT)
  3. Other muscarinic antagonists or botulinum toxin (if taken during the last 2 weeks prior to week 12/EoT assessment);

  In all cases, a hypothetical strategy will be applied.
- Population-based summary: Change from baseline to week 12/EoT in micturition frequency/day (= primary endpoint, primary estimator), compared to placebo.

For this placebo-controlled study, this "de jure" estimand is considered the appropriate choice. As some effect is expected early in the study and with continued treatment could be expected to be maintained, it is considered appropriate to impute for subjects with missing values at visit 7/week 12/EoT by their last recorded post baseline visit value prior to it. This strategy also assumes that the responses after the intercurrent event would be same as for those patients continuing on their respective treatment.

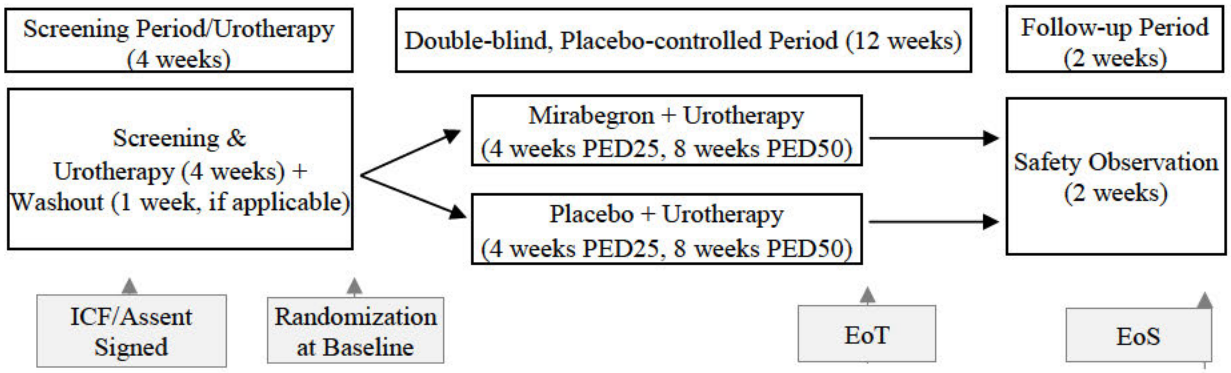
2.2 Study Design

This is a double-blind, randomized, multicenter, parallel group, placebo-controlled sequential dose titration study to evaluate efficacy, safety and pharmacokinetics of mirabegron in pediatric subjects with OAB. Male and female pediatric subjects 5 to <18 years of age with OAB; as defined according to the International Children's Continence Society (ICCS) will receive 4 weeks of urotherapy prior to randomization.

Planned total number of study sites included approximately 65 study sites across Europe, Latin America, Africa, Middle East, Asia-Pacific and North America.

Figure 1 presents the study schema.

Figure 1 Study Schema



EoS: End of Study; EoT: End of Treatment; ICF: informed consent form;
PED25: pediatric equivalent dose 25 mg; PED50: pediatric equivalent dose 50 mg

The study consists of 3 periods with a total duration of 18 weeks.

- Screening period/urotherapy (4 weeks):
  - This period starts with visit 1/week -4 (screening) and ends with visit 3/week 0 (baseline). After informed consent/assent has been obtained and immediately after eligibility has been confirmed at visit 1/week -4 (screening), subjects using prohibited medication will complete 1 week of washout (if applicable), while beginning 4 weeks of urotherapy.
  - After a successful screening visit (visit 1/week -4 [screening]), all subjects will need to complete a 2-day bladder e-diary (weekend) to get acquainted with the bladder e-diary and the assessments. Completion of this bladder e-diary should start in the weekend prior to visit 2. All subjects will also complete a 7-day bladder e-diary the week prior to the baseline visit. The 7-day diary will consist of a 5-day weekday bladder e-diary and 2-day weekend e-diary.
- Double-blind, placebo-controlled period (12 weeks):
  - This period starts with the day after visit 3/week 0 (baseline) and ends with visit 7/week 12 (EoT).
  - At visit 3/week 0 (baseline) inclusion and exclusion criteria will be evaluated. Subjects continuing urotherapy who still meet the OAB entry criteria at baseline will be randomized. Subjects whose symptoms are not satisfactorily controlled with urotherapy and still fulfill the inclusion/exclusion criteria will enter the study. These subjects will be randomized to receive mirabegron in PED25 or placebo using a 1:1 ratio. Subjects with a body weight of $\geq$ 35 kg are to receive the tablet unless unable to swallow tablets and would be provided the oral suspension as an alternative. Subjects with a body weight < 35 kg or those who cannot be dosed with the tablet will receive an oral suspension. Daily investigational product (IP) administration will start on day 1 (i.e., the day after this visit) and continue at this dose until visit 5/week 4 (i.e., for 4 weeks). Urotherapy will continue throughout the study treatment period until visit 7/week 12 (EoT).

- At visit 5/week 4, dose up-titration to mirabegron in PED50 will be performed unless the investigator determines that the subject is adequately treated for OAB at the PED25 dose or there are safety concerns identified and considered associated with the use of PED25. Dose down-titration from PED50 to PED25 can be done at any time thereafter for safety reasons.
- Subjects will start with the subsequent 7-day bladder e-diaries approximately 7 days prior to the indicated visit (or TC).
- Pharmacokinetic blood samples will be collected at visit 5/week 4 and visit 7/week 12 (EoT) as indicated in the Schedule of Assessments, Table 1 in protocol

- Follow-up period (2 weeks):
  - This period starts the day after visit 7/week 12 (EoT) and ends with visit 8/week 14 (end of study [EoS]). The follow-up period is applicable to all subjects who have been randomized and received IP.
  - At visit 7/week 12 (EoT), IP administration will be stopped and a safety observation period of 2 weeks will start.

Details of the schedule of clinical assessments are available in the protocol.

An independent DSMB will be established. A separate charter will describe the responsibilities of the DSMB.

A comparative interim analysis will be performed after 50% of children planned to be randomized have had their visit 7/week 12/EoT assessment. The interim analysis will determine if the chance of a positive study with respect to the primary endpoint at the end of the study/Final Analysis is high enough to justify continuation of the study; otherwise, the study will be stopped for futility. For details see Section 6.9.

2.3 Randomization

Subjects will be entered into the interactive response technology (IRT) system at screening and assigned a subject number. At Visit 3 (baseline) randomization will be performed via the IRT system and treatment assigned in a 1:1 ratio to mirabegron or placebo. Prior to the initiation of the study treatment, on day 1, the site staff will contact the IRT system in order to determine the randomly assigned treatment. Subjects will be stratified by age category: 5 – <12 years and 12 – <18 years of age.
Subjects will be assigned a subject number at study entry (i.e., signing the ICF/assent). The subject numbers will be sequential and rising. The subject number will comprise of 10 digits; 5 digits for the site number (provided by the sponsor) and 5 digits for the consecutive subject number.
Subjects will be randomized to receive IP according to the randomization schedule obtained via the IRT system. The study site personnel will dispense the treatment according to the IRT system's assignment. Specific procedures for randomization through the IRT system are contained in the study procedures manual.
An enrolled subject who withdraws or discontinues before dosing or before randomization will be considered a screen failure. Randomized subjects who withdraw or discontinue will not be replaced.

3 SAMPLE SIZE

Efficacy of mirabegron in children will be based on both primary and secondary endpoints. Given the clinical hurdle to demonstrate efficacy in a population challenging to recruit, an alpha-level of 10% would appear reasonable and justified. In a study with propiverine in children, Marschall-Kehrel (2009) showed a difference in mean number of micturition episodes/24 hours of 0.8 between propiverine and placebo, with a standard deviation (SD) between 2.2 and 2.3. Assuming at least similar efficacy of mirabegron, with a treatment effect of 0.9 micturitions per day between mirabegron and placebo and a common SD of 2.3, a sample size of 82 evaluable subjects (children) per treatment group would provide a power of 80%.

These assumptions, together with levels for Type-1 error and power were agreed upon with Pediatric Committee in the current Pediatric Investigation Plan.

Under the assumption that after randomization about 10% of the subjects will not be evaluable for the analysis of the primary endpoint (i.e., not fulfilling the criteria for inclusion in FAS), at least 184 children (5 to < 12 years of age) must be randomized in order to have at least 164 evaluable children for the analysis of the primary efficacy endpoint.
For adolescents, no formal sample size will be calculated, at least 32 adolescents (12 to <18 years of age) must be randomized, yielding at least 16 adolescents on mirabegron.

Further assuming a 50% screen failure rate, approximately 368 children have to be enrolled to achieve 184 children (5 to <12 years of age) randomized and approximately 64 adolescents will be enrolled to achieve at least 32 adolescents (12 to <18 years of age) randomized (at least 92 children and 16 adolescents must be randomized to mirabegron).

Recruitment will continue until the minimum number in both children and adolescents has been achieved.

4 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the following analysis sets will be used for the analyses.

The determination of whether subjects are included or excluded from the safety and efficacy analysis sets will be made prior to database hard-lock for the primary report and prior to unblinding.

4.1 All Screened Set

The All Screened Set consists of all subjects for whom a valid informed consent/assent is available, as per applicable local law.

Consent/assent should be obtained from the subject and his/her parent(s)/legal guardian(s), before start of pre-investigational period. Up to three signatures might be available from the informed consent/assent. For calculations using the date of informed consent/assent the last available date will be used.

The All Screened Set will be used to summarize disposition of subjects who were screened.

4.2 All Randomized Set

The All Randomized Set consists of all randomized subjects, i.e. all those subjects with a randomization number.

The All Randomized Set will be used to summarize disposition of subjects who were randomized to treatment.

4.3 Full Analysis Set

The Full Analysis Set (FAS) consists of all subjects who are randomized and receive at least 1 dose of IP and have at least 1 post baseline measurement for mean number of micturitions per 24 hours. The FAS will be analyzed by treatment arm as randomized (i.e., treatment arm based on randomization assignment).

This will be the primary analysis set for efficacy analyses. Selected demographic and baseline characteristics will also be summarized for the FAS.

4.4 Safety Analysis Set

The Safety Analysis Set (SAF) consists of all subjects who took at least 1 dose of IP. The safety set will be analyzed by treatment arm as treated (i.e., based on the treatment the subject actually received rather than the treatment to which the subject was randomized).

The SAF will be used for summaries of demographic and baseline characteristics and all safety- and tolerability-related variables.

4.5 Per Protocol Set

The Per Protocol Set (PPS) consists of all subjects who are randomized and receive at least 1 dose of IP and have at least 1 post baseline measurement for mean number of micturitions per 24 hours, and who do not meet criteria for PPS exclusion listed in Section 4.5.1 of this

SAP. Final judgments on exclusion of subjects from the PPS are to be made prior to data base hard lock and will be documented in the Classification Specification document.

The PPS will be used for sensitivity analyses of the efficacy data. Selected demographic and baseline characteristics may also be summarized for the PPS.

4.5.1 Reasons for Exclusion from Per Protocol Set

A subject may be excluded from the PPS if there has been a deviation from the protocol sufficient to affect the assessment of the efficacy of the study drug. Such deviations may include (but are not limited to) ineligible subjects being enrolled in the study, interventions performed that were prohibited by protocol, or when subjects did not adhere adequately to the study treatment or outcome assessments.

There will be no partial exclusion (e.g., of particular timepoints only) of a subject from the PPS; all of a subject's data will be excluded.

Some reasons for exclusion of a subject from the PPS are given below: however the decision on whether a particular protocol deviation (see Section 6.2.2) requires the exclusion of a subject will be confirmed at the Analysis Set Classification Meeting (ASCM) with all reasons for the decision documented in the meeting minutes.

Eligibility Deviations

The inclusion and exclusion criteria for the study are detailed in Sections 5.1 and 5.2 of the protocol. These are all assessed at Screening (Visit 1) and selected criteria are also assessed at baseline (Visit 3).

A subject will be excluded from the PPS if:

- Any of inclusion criteria 2 and 15 are not met
- Any of exclusion criteria 2, 4, 8, 10 and 28 are met

Violations of other inclusion and exclusion criteria may also result in exclusion of the PPS if it is considered that there is a risk that the violation affects the assessment of the efficacy of the study drug. All such violations will be considered on an individual basis at the ASCM.

Prohibited and restricted medications and non-drug therapies

A subject will be excluded from the PPS if they received prohibited medications or non-drug therapies. These include, but are not limited to, the following (see Table 12.6 in protocol for a more complete list):

- Anti-muscarinics (including tricyclic and tetracyclic antidepressants), desmopressin, alpha-blockers and prescribed or OTC drugs that are potent CYP3A4 inhibitors (e.g., ketoconazole). They have to be stopped at screening (Visit 1) as soon as informed consent/assent has been obtained and eligibility has been confirmed and are prohibited during the study
- A history of intravesical treatment with botulinum toxin within 9 months prior to screening (Visit 1) and intravesical administration of botulinum toxin during the study is prohibited

- A history of electrostimulation or bladder training within 2 weeks prior to screening (Visit 1) and at any time during the study (outside of study related urotherapy) is prohibited.

Administration of restricted medications will be assessed during medical review to determine if the usage has fallen outside the protocol restrictions and if there is a risk that baseline or endpoint measurements may be affected. A list of restricted medications is given in Appendix 12.6 of the protocol.

Compliance to study medication

A subject will be excluded from the PPS for any of the following reasons based on compliance to the study medication (see Section 5.5.2 for details on the calculation of the compliance rate between study visits and over the whole double-blind study period):

- Subject took double-blind study medication for less than 70 days.
- Compliance with the double-blind study medication during the entire double-blind treatment period is less than 70%.
- Compliance with the double-blind study medication has been assessed as being less than 80% between visit 6/week 8 and visit 7/week 12/EoT.

All evidence and reasons for determining a violation based on compliance to the study medication will be documented in the minutes of the ASCM.

Error in study drug administration

A subject will be excluded from the PPS if they were dispensed incorrectly allocated study medication at any time during the double-blind treatment period.

4.6 Pharmacokinetics Analysis Set

The Pharmacokinetic Analysis Set (PKAS) will consist of all subjects who took at least 1 dose of IP and contributed at least one pharmacokinetic sample in which the date and time of the sample and prior dose are known. Any formal definitions for exclusion of subjects or time points from the PKAS will be documented in the classification specifications and be determined at the classification meeting.

Since the actual bioanalytical results may only become available after the data review meeting, additional data points may be excluded at the time of pharmacokinetic analysis at the discretion of the GCMSL. These data points will be documented in the modeling report.

The PKAS is used for all tables and graphical summaries of the PK data.

4.7 Pharmacodynamic Analysis Set

Not applicable.

5 EFFICACY AND SAFETY ENDPOINTS

Data for the efficacy endpoints will be recorded on an e-diary. This e-diary consists of a:

- 5-day Weekday Diary (bladder diary, Monday – Friday), which is used for a 5-day period to record micturition frequency and incontinence episodes, and
- 2-day Weekend Episodic Diary (Saturday and Sunday), which is used to record micturition frequency, incontinence episodes, (pee) volume measurements, and worst leakage severity.

A valid diary day for:

- "Volumes"
  In the 2-day Weekend Episodic Diary, is any e-diary day for which at least one "Pee Volume" greater than 0 mL is recorded with complete date and time.
- "Micturitions"
  - In 2-day Weekend Episodic Diary, is any diary day for which at least once "Pee Volume" greater than 0 is recorded.
  - In 5-day Week Diary, is any diary day for which the response to the question "Number of times using the toilet during the day/night" is not missing.

- "Incontinence episodes"
  - In 2-day Weekend Episodic Diary, is any diary day for which at least:
    - One micturition ("Pee Volume" >0) and
    - One "Leakage" or "Pee in toilet and Leakage" are recorded.
  - In 5-day Week Diary, is any diary day for which:
    - The response to the question "How was your day/night" is not missing.

A valid diary for an analysis visit is a diary with at least 4 valid diary days within that analysis visit window. If for more than 7 days the diary is filled in prior to a visit, only the last 7 days prior to that visit will be used for calculation of the efficacy endpoints.

The number of valid diary days: Count of the valid diary days during the 7-day diary period. Days of visits to the clinic will not be counted as valid diary days.

In the event that in the 2-day Weekend Episodic Diary there are two entries with the same collection date and time but with different times of entry on the device, the last entry (time) will be used for analysis.

For several of the secondary efficacy endpoints, diary data will be classified as either "daytime" or "nighttime", where "daytime" is considered to mean "between waking up in the morning and going to sleep later the same day or next day" and "nighttime" means "between going to sleep on a day and waking up on the same or next day".
Whether an entry into the 5-day Week Diary is done during daytime or nighttime is captured on this diary.

For entries into the 2-day Weekend Episodic Diary, "daytime" and "nighttime" will be derived from the time the subject went to sleep and the time the subject awakened. These times are entered on this diary by the subject.

The go-to-bed time for Sunday is not captured in the diary. In order to distinguish between daytime and nighttime on Sunday the go-to-sleep time will be imputed as occurring a short time after the second (evening) SBPM was performed: time of second SBPM plus 1 hour. If there is no time entered in the diary for the evening SBPM, then the go-to-sleep time will be imputed as 23:59 (on Sunday).

5.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to week 12/EoT in mean number of micturitions per 24 hours in children (5 to <12 years of age).

Number of micturitions per 24 hours is calculated from:

- The response to the questions "Number of times using the toilet during the night" and "Number of times using the toilet during the day" on the 5-day Week Diary,

and

- The number of times a "Pee in toilet" or a "Pee in toilet and leakage" was entered into the 2-day Weekend Episodic Diary.

For a weekend day:

- The daytime micturitions will be derived from the number of times a "Pee in Toilet" or a "Pee in Toilet and Leakage" was entered into the 2-day Weekend Episodic Diary between the time the subject woke up (exclusive) and went to sleep (inclusive).
- The nighttime micturitions will be derived from the number of times a "Pee in Toilet" or a "Pee in Toilet and Leakage" was entered into the 2-day Weekend Episodic Diary between the time the subject went to sleep (exclusive) and woke up (inclusive).

The total number of micturitions per weekend day is equal to the total number of times, in the diary, an amount of pee was recorded during daytime and nighttime for that day.

For each subject, the mean number of micturitions at Visit x (x=4, 5, 6, 7), using all available (non-missing) number of micturitions per 24 hours for the valid days in the diary (at least 4 and maximally 7), will be calculated as:

$$\frac{\text{Sum of the number of micturitions (per 24 hours) over the valid diary days prior to Visit x}}{\text{Number of valid diary days}}$$

Note:

- If for a week day (Monday – Friday):
  - the number of daytime micturitions is missing but the number of nighttime micturitions is not missing then the number of micturitions for that day is considered to be the number of micturitions during the nighttime.

- the number of nighttime micturitions is missing but the number of daytime micturitions is not missing then the number of micturitions for that day is considered to be the number of micturitions during the daytime.
- both daytime and nighttime micturitions are missing then the number of micturitions for that day is considered missing.
    - If for a weekend day (Saturday – Sunday):
- No pee volume is recorded for Saturday or for Sunday then the number of micturitions is missing for Saturday or Sunday, respectively.
- No pee volume is recorded for Saturday and for Sunday then the number of micturitions is missing for both days.

If the wake-up time is the same as the time of micturition, then the micturition will be counted as a nighttime micturition.
Note: the first micturition occurring at or after the time of wake-up is considered a nighttime micturition.

A micturition episode is counted as micturition regardless of whether volume voided was recorded or not.
A micturition will also be counted if the subject assessed an episode as both micturition and incontinence (leakage).

If data of more than 7 days prior to a visit is available, only the last 7 days prior to the visit will be used for the analysis.

The baseline will be defined by the micturitions results obtained from the diary 7 days prior to visit 3/week 0 (baseline).

The visit 7/week 12/EoT visit will be defined as the subject's value at week 12 or in case of missing week 12 data, the last available result post-baseline.

Subjects without baseline or any post-baseline value will not be included in the change from baseline calculation although they will be included in the summary for either baseline or any post-baseline visit whichever timepoint the data is available.

Change from baseline to visit 7/week 12/EoT is defined as:
Result at visit 7/week 12/EoT – Result at baseline (visit 3/week 0).

5.2 Secondary Efficacy Endpoints

The following secondary efficacy endpoints, as derived from the e-diary, are the change from baseline to the end of the 12-week treatment period in children (5 to <12 years of age) in:

- Mean volume voided per 24 hours
- Maximum volume voided
- Mean number of daytime incontinence episodes per 24 hours
- Mean number of nighttime incontinence episodes per 24 hours
- Mean number of daytime micturitions per 24 hours
- Number of dry (incontinence free) days per 7 days

For all above secondary efficacy endpoints the value calculated from the e-diary period 7 days before visit 3/week 0/baseline will be regarded as the baseline value.

5.2.1 Mean Volume Voided per 24 hours

Mean volume voided is derived from "Pee Volume" of the 2-day Weekend Episodic Diary.

For each subject, the mean volume voided per day will be calculated as the sum of the volumes voided on that (valid diary) day divided by the number of times a volume was recorded on that day in the 2-day Weekend Episodic Diary.
The mean volume voided at Visit x (x=4, 5, 6, and 7) will be calculated as the mean value of the mean volumes voided on each of the two days
If volumes are recorded on 1 single (valid) day of the 2-day Weekend Episodic Diary, then the mean volume will be the sum of all available volumes of that day divided by the number of times a volume was recorded.
If no volumes are recorded on either (valid) measuring day, then the mean volume will be missing.

5.2.2 Maximum Volume Voided

Maximum volume voided data will be derived from "Pee Volume" of the 2-day Weekend Episodic Diary.

For each subject, the maximum volume voided at Visit x (x=4, 5, 6, and 7) is the largest (non-zero) volume recorded over both of the 2 (valid) measuring days in the diary. If volumes are recorded on 1 single (valid) day of the weekend diary, then the maximum volume voided per day is calculated as the largest of all available (non-zero) voided volumes recorded on that day.
If no volumes are recorded on either (valid) measuring day, then maximum volume voided will be missing.

5.2.3 Mean Number of Daytime Incontinence Episodes per 24 hours

For a week day the number of Incontinence Episodes per 24 hours will be derived from "Number of Leakages During the Day" and "Number of Leakages During the Night" of the 5-day Week Diary. Incontinence episodes on measuring days will be classified as daytime or nighttime episodes as entered in the 5-day Week Diary.

For a weekend day the number of incontinence episodes will be derived from the times a "Leakage" or a "Pee in Toilet and Leakage" was entered into the 2-day Weekend Episodic Diary. Whether an Incontinence Episode was during daytime or nighttime will be based on the time-points the subject went to sleep and woke-up.

For each subject, the mean number of incontinence episodes per day/night (during the day/night) at Visit x (x=4, 5, 6, and 7) will be calculated using all available (non-missing) number of incontinence episodes for the valid diary days in the 5-day Week Diary and in the 2-day Weekend Episodic Diary for daytime in the 7-day period before the visit:

$$\text{Mean No. of Daytime Incontinence Episodes at Visit x} =$$

$$\frac{\text{Total number of "Daytime Leakage Counts" over the Valid Diary Days prior to Visit x}}{\text{Number of Valid Diary Days}}$$

5.2.4 Mean Number of Daytime Micturitions per 24 hours

For a week day the daytime micturitions will be derived from "Number of Times using the Toilet During the Day" was entered into the 5-day Week Diary.

For a weekend day the daytime micturitions will be derived from the number of times a "Pee in Toilet" or a "Pee in Toilet and Leakage" was entered into the 2-day Weekend Episodic Diary between the time the subject woke-up (exclusive) and went to sleep (inclusive). However, the first micturition occurring at or after the time of wake-up it is considered a nighttime micturition, see Section 5.1.

The total number of micturitions per weekend day is equal to the total number of times, in the diary, an amount of pee was recorded during daytime for that day.

For each subject, the mean number of daytime micturitions at Visit x (x=4, 5, 6, and 7) in the week before the visit will be calculated as:

$$\frac{\text{Sum of the Number of Daytime Micturitions (per day) over the Valid Diary Days prior to Visit x}}{\text{Number of Valid Diary Days}}$$

5.2.5 Number of Dry (incontinence free) Days per 7 Days at the End of the 12-weeks Treatment Period

For a week day a "Dry (incontinence free) Day" is defined as a day where the response is "Dry" to the question "How was your Day" and to "How was your Night".
Let D_{dry} be the number of valid diary days where the response to both questions was "Dry".
Let D_{wet} be the number of valid diary days where the response to one of the two questions or to both questions was "Wet".

For a weekend day a "Dry (incontinence free) Day" is defined as a day where no "Leakage" and no "Pee in toilet and leakage" are reported. A "Wet (incontinence) Day" is defined as a day where a "Leakage" or a "Pee in toilet and leakage" is reported.

If $(D_{dry} + D_{wet}) > 3$, then the number of dry days per 7 days at Visit 7 will be calculated for a subject as:

$\frac{D_{dry}}{(D_{dry}+D_{wet})} \times 7$, otherwise its value is missing.

5.3 Exploratory Endpoints

The following exploratory endpoints, as derived from data collected in the e-diary, are:

1. Percentage of subjects with a reduction in daytime incontinence episodes
2. Improvement from baseline in worst incontinence grading
3. Mean number of micturitions per 24 hours adjusted for fluid intake
4. Mean volume voided peer micturition as a percentage of EBC
5. Categorized mean volume voided per micturitions a percentage of EBC
6. Mean number of micturitions per 24 hours (adolescents only)
7. Mean number of daytime micturitions per 24 hours (adolescents only)
8. Mean volume voided per 24 hours (adolescents only)
9. Maximum volume voided (adolescents only)
10. Mean number of incontinence episodes per 24 hours (adolescents only)
11. Number of dry (incontinence-free) days per 7 days (adolescents only)
12. Mean number of daytime grade 3 or 4 (PPIUS) urgency episodes per 24 hours (adolescents only)

5.3.1 Percentage of Subjects With a Reduction in Daytime Incontinence Episodes (< 50% reduction [non-responder], 50% [partial responder] and 100% [responder])

Number of daytime incontinence episodes per 24 hours will be derived from:

- "Number of leakages during the day" entered into the 5-day Weekday Diary, and
- number of times "Leakage" or "Pee in toilet and leakage" was entered into the 2-day Weekend Episodic Diary between awake time and sleep time.

For each subject, the percent change from baseline to a Visit x (x=4, 5, 6 and 7) in the number of daytime incontinence episodes will be calculated as follows:

$$R_x = \frac{\text{No. Daytime Incontinence Episodes at Visit x} - \text{No. Daytime Incontinence Episodes at Baseline}}{\text{No. Daytime Incontinence Episodes at Baseline}} \times 100\%,$$

where the number of daytime incontinence episodes at Visit x is equal to the sum of the number of incontinence episodes recorded over the 7-day period prior to Visit x. Same calculation for baseline.
If the number of daytime incontinence episodes at baseline is equal to zero, a constant equal to 0.5 will be added to the baseline number of leakages to allow denominator calculation.
If the number of daytime incontinence episodes at baseline and/or at a post-baseline visit is missing then the response category in respect to incontinence will be missing.

Definition of responder:

- Complete responder will be defined as a subject with a 100% improvement from baseline (i.e. R_X= -100%).
- Partial responder will be defined as a subject with a percent improvement from baseline ≥50% and <100% (i.e. -100% < R_X ≤ -50%).
- Non-responder will be defined as a subject with an improvement from baseline <50% or a worsening from baseline (i.e. R_X > -50%).

5.3.2 Improvement from Baseline in Worst Incontinence Grading

Worst Incontinence Grading is derived from:

- The response to "Worst Leakage during the Day" and "Worst Leakage during the Night" from the 5-day Week Diary, and is equal to the worst outcome from both daytime and nighttime responses during a diary day.
- The responses to "Worst Leakage Severity" every time it is reported in the 2-day Weekend Episodic Diary. The worst response to the questions given during a diary day will be used for analysis.

Possible outcomes are: Dry; Slightly Wet; Moderately Wet; and Fully Wet.
Dry will be derived from the question on the 5-day Week Diary: "How was your day/night", and from the 2-day Episodic Weekend Diary if no leakages are reported during a diary day.

For each subject and visit the worst incontinence grade will be determined as the worst grading over the valid diary days preceding the study visit.

Improvement:
For Visit x (x=3, 4, 5, 6 and 7) the average over the valid week and weekend diary days will be calculated from the score of what was the worst leakage during the day", *i.e.,* 0 - Dry, 1 - Slightly Wet; 2 - Moderately Wet; and 3 - Fully Wet:

$$\text{Worst Incontinence Grade at Visit x} = \frac{\text{Sum of Worst Incontinence Gradings over the Valid Diary Days prior to Visit x}}{\text{Number of Valid Diary Days}}$$

For each subject, improvement at Visit x if:
Worst Incontinence Grade (Visit x) - Worst Incontinence Grade (Visit 3 [Baseline]) < 0.

5.3.3 Mean Volume Voided as a Percentage of Expected Bladder Capacity

To calibrate the mean volume voided with bladder size, the ratio of mean volume voided to the subject's EBC (= [age+1] × 30 in mL for children and 400 mL for adolescents) will be calculated at baseline and post-baseline study visits and expressed as a percentage. These ratios will be summarized together with the change from baseline by age group and overall.

The following variables will be summarized, for each age group and overall, by treatment group at each visit using number and percentage of subjects in the following categories:

- Ratio of mean volume voided to EBC (%): ≤50%, >50% to ≤75%, >75% to ≤100%, >100%

- Change from baseline in ratio of mean volume voided to EBC (%): <0%, 0% to <15%, 15% to <30%, ≥30%

5.3.4 Mean Number of Micturitions per 24 hours, Adjusted for Fluid Intake

For this endpoint the voided volumes will be taken as surrogate for the fluid intake.

In Sender Herschorn *et al.*, (2017) page 56 Materials and Methods, the following subdivision is presented of the number of micturitions at a visit, say visit x.

Number of Micturitions at Visit x (x= 4, 5, 6, and 7) =

Total Volume Voided (Visit x) / Mean Volume Voided (Visit x) =

[Total Volume Voided (Baseline) + ΔTotal Volume Voided]/Mean Volume Voided (Visit x) =

Total Volume Voided(Baseline) / Mean Volume Voided(Visit x) +

ΔTotal Volume Voided / Mean Volume Voided(Visit x),

where Δ is the difference in total volume voided between Visit x and Baseline [=Total Volume Voided (Visit x) – Total Volume Voided (Baseline)].

This can be viewed as a partition of Number of Micturitions at Visit x into 2 parts as follows:

- Total Volume Voided (Baseline) / Mean Volume Voided (Visit x) is the number of micturitions per 24 hours that would be required at Visit x to void the total daily volume, if this total volume remains unchanged from baseline, that is, if treatment did not affect subjects' fluid intake.
- Δ Total Volume Voided / Mean Volume Voided(Visit x) is the additional number of micturitions per 24 hours (vs baseline) required to void the extra fluid intake.

For each subject, the Mean Change in Number of Micturitions per 24 hours at Visit x (x= 4, 5, 6, and 7), adjusted for fluid intake will be calculated as:

$$\frac{\text{Total Volume Voided at Visit x} - \text{Total Volume Voided at Baseline}}{\text{Mean Volume Voided per Micturition at Visit x}}$$

Example:

A subject voided the following volumes:

- At baseline,
  Saturday: 200 ml, 200 ml, 200 ml, 200 ml and 200 ml, and
  Sunday: 200 ml, 200 ml, 200 ml and 200 ml
- At visit x,
  Saturday: 250 ml, 250 ml, 250 ml and 250 ml, and
  Sunday: 250 ml, 250 ml and 250 ml

This results in:
Total Volumes Voided per 24 hours at baseline and at visit x are 900 ml and 875 ml, respectively. Mean Volume Voided per 24 hours at visit x is 250 ml,

Total Volume Voided (baseline) / Mean Volume Voided (visit x) = 900 ml / 250 ml = 3.6, which means that 3.6 micturitions per 24 hours would be required at visit x to void the same total volume as voided at baseline (3.6 x 250 = 900 ml). This is the expected mean number of micturitions per 24 hours at visit x.

Mean number of micturitions at baseline and at visit x are 4.5 and 3.5, respectively. Mean difference in the number of micturitions per 24 hours adjusted for fluid intake compared to expected mean number of micturitions per 24 hours at visit x is (875-900)/250 = -0.1.

5.3.5 Mean Number of Micturitions per 24 hours (adolescents only)

See Section 5.1.

5.3.6 Mean Volume Voided per 24 hours (adolescents only)

See Section 5.2.1.

5.3.7 Maximum Volume Voided (adolescents only)

See Section 5.2.2.

5.3.8 Mean Number of Incontinence Episodes per 24 hours (adolescents)

See Section 5.2.3.

5.3.9 Mean Number of Daytime Micturitions per 24 hours (adolescents only)

See Section 5.2.4.

5.3.10 Number of Dry (incontinence-free) Days per 7 Days at the End of the 12-weeks Treatment Period (adolescents only)

See Section 5.2.5.

5.3.11 Mean Number of Daytime Grade 3 or 4 (PPIUS) Urgency Episodes per 24 hours (adolescents only)

Result is obtained from "PPIUS-Urgency of the Micturition" from 2-day Weekend Episodic Diary.

Scoring PPIUS scale:

- Grade 0: No urgency – I felt no need to empty my bladder, but did so for other reasons
- Grade 1: Mild urgency – I could postpone voiding as long as necessary, without fear of wetting myself
- Grade 2: Moderate urgency – I could postpone voiding for a short while, without fear of wetting myself
- Grade 3: Severe urgency – I could not postpone voiding, but had to rush to the toilet in order not to wet myself
- Grade 4: Urge incontinence – I leaked before reaching the toilet

Let N_{34} equal the number of times that an adolescent subject recorded a grade 3 or 4 urgency on a diary day. If there are no grade 3 or 4 urgency episodes on a valid diary day, then N_{34} equals 0.

For each subject, the mean number of grade 3 or 4 urgency episodes per 24 hours, at Visit x (x=4, 5, 6, and 7), is the mean of N_{34} over the valid diary days of the 2-day Weekend Episodic Diary.

5.4 Safety Endpoints

Safety will be assessed by evaluation of the following variables:

- Treatment-emergent adverse events (TEAEs; frequency, severity, seriousness, and relationship to study drug)
- Clinical laboratory variables (hematology, biochemistry including liver enzymes and total bilirubin, and urinalysis)
- Vital signs (systolic and diastolic blood pressure and pulse rate)
- 12-lead electrocardiogram (ECG)
- Post Void Residual (PVR) Volume
- Acceptability and palatability questionnaires

5.4.1 Adverse Events

AEs will be coded using MedDRA. A TEAE is defined as an AE observed after starting administration of the IP until 30 days after the final administration of IP.
If the adverse event occurs on Day 1 and the onset check box on the (AE) eCRF is marked "Onset after first dose of study drug" or the onset check box is left blank, then the adverse event will be considered treatment emergent. If the adverse event occurs on Day 1 and the onset check box is marked "Onset before first dose of study drug", then the adverse event will not be considered treatment emergent. If a subject experiences an event both during the pre-investigational period and during the investigational period, the event will be considered as TEAE only if it has worsened in severity (i.e., it is reported with a new start date).

A drug-related TEAE is defined as any TEAE with at least a possible relationship to study treatment as assessed by the investigator or with missing assessment of the causal relationship.

An AE designated as serious by either the investigator or determined by Important Medical Event process will be summarized as an SAE. However, if the event is on IME list but medical monitor agrees with investigator that event is not SAE, then event can be downgraded.

Common TEAEs are defined as preferred terms (PTs) that have been reported by at least 5% of the subjects in any of the treatment groups.

When an AE start or stop date is missing, the date will be imputed using the rules in Section 6.10.2.2.

Adverse Events of Special Interest

Adverse Events of Special Interest include:

- Abnormal cardiac electrophysiology (including QTc prolongation), cardiac arrhythmia, and other cardiovascular adverse events

- Increased blood pressure as assessed through vital signs monitoring
- Hypersensitivity reactions
- Urinary retention
- Abnormal nervous system events: seizure, syncope, headache and dizziness

AEs of interest will be identified using all Preferred Terms (PT) from Standardized MedDRA queries (SMQ), see Table 2, or sponsor-defined list of search terms, see Appendix 9.2.
Note: Most recent MedDRA version in which the AEs are coded will be used.

Table 2 Standardized MedDRA Queries

CV - Increased blood pressure	Hypertension SMQ – Narrow search
CV - QT prolongation	Torsade de pointes/QT prolongation SMQ – Broad search
CV - cardiac arrhythmia, cardiovascular adverse events	Arrhythmia related investigations signs and symptoms SMQ – Broad search, Supraventricular tachyarrhythmias (SMQ) – Broad search, Tachyarrhythmia terms nonspecific (SMQ) – Narrow search, Ventricular tachyarrhythmias (SMQ) -narrow search plus ventricular tachycardia
Hypersensitivity reactions	Hypersensitivity SMQ – Narrow search
Seizure	Lower Level Term 10039906
Syncope	Syncope SMQ – Narrow search (also included as part of the CV - Increased heart rate, tachycardia, atrial fibrillation, and palpitations) PT 10042772
Headache	Headache PT 10019211
Dizziness	Dizziness PT 10013573

5.4.2 Vital Signs

5.4.2.1 Clinical Measurement of Vital Signs

Single blood pressure, pulse and body temperature measurements will be performed at visit 1/screening, visit 3/baseline, visit 5/week 4, visit 7/week 12 (EoT), and visit 8/week 14 (EoS). Clinic measurements of vital signs at visit 1/screening and visit 3/baseline will be used to assess eligibility.

The preferred method of blood pressure measurement is via the auscultatory technique. If this method is not available at the study site, measurements will be per standard clinic practices.

Single measurements of pulse and body temperature will be measured in the sitting position (when possible, otherwise supine, but always in the same position).

For subjects enrolled under:

- Protocol version 1 and 2, one single blood pressure measurement will be taken.

- Protocol version 3 or higher (if applicable), at least two blood pressure measurements will be taken separated by 5 minutes and the average of these measurements will be the documented blood pressure.

5.4.2.2 Self-Measurement of Vital Signs

Self-blood pressure monitoring (SBPM; SBP and DBP) will be measured at visit 2/week 2, visit 3/baseline, visit 4/week 2, visit 5/week 4, visit 6/week 8, and visit 7/week 12 (EoT) and whenever deemed necessary by the investigator.

SBPM will be measured once in the morning and evening during the 2-day Weekend Episodic Diary collection period. Measurements will be taken in the sitting position (when possible, otherwise supine, but always in the same position). Preferably, the right arm should be used. Morning measurement should be taken before IP intake and evening measurement should be taken prior to bedtime.

5.4.2.3 Conversion of Blood Pressure to Percentiles

Both the systolic and diastolic blood pressure site measurements will additionally be converted to percentiles specific to the age (years), sex and height (cm) of the subject using the following steps [NIH, 2005]:

- The height of a subject is measured at screening and baseline. Height at baseline will be used.
- Let Z = the z-score (i.e. the number of standard deviations above or below the mean) of the height (cm) of the subject at the study visit relative to healthy subjects of the same age and gender using CDC growth charts [CDC, 2000]. The SAS code to calculate this is located on the Centers for Disease Control (CDC) and Prevention website: https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm.
- Let BP = the blood pressure (mmHg) that is being converted (either systolic or diastolic) and Y=the age of the child in years.
- Taking the coefficients from the appropriate column in Table 3 below, compute Z_{BP}, the Z-score for the blood pressure:

$$Z_{BP} = \left[BP - \alpha - \sum_{j=1}^{4} \beta_j (Y-10)^j - \sum_{k=1}^{4} \gamma_k Z^k \right] \Big/ \sigma$$

- If Φ(.) is the area under the standard normal curve to the left of Z, calculate P, the gender, age and height specific percentile of the blood pressure as $P=\Phi(Z_{BP})*100\%$.
- The exact age at a visit will be calculated as:

$$Age = \frac{(D_V - D_B + 1)}{365.25}$$

where D_v is the visit date and D_B is the subject's date of birth. When the exact age at a visit is missing and cannot be calculated because of a missing date of birth, but where the

age entered at screening is known only to be Y years (an integer), the percentile will be calculated for age=Y+0.5.

Table 3 Coefficients for Conversion of Blood Pressure to Percentiles

Variable		Symbol	Systolic BP		Diastolic BP	
			Male	Female	Male	Female
Intercept		A	102.19768	102.01027	61.01217	60.50510
Age	(Age-10)	β_1	1.82416	1.94397	0.68314	1.01301
	$(Age-10)^2$	β_2	0.12776	0.00598	-0.09835	0.01157
	$(Age-10)^3$	β_3	0.00249	-0.00789	0.01711	0.00424
	$(Age-10)^4$	β_4	-0.00135	-0.00059	0.00045	-0.00137
Height	Z	γ_1	2.73157	2.03526	1.46993	1.16641
Z-Score	Z^2	γ_2	-0.19618	0.02534	-0.07849	0.12795
	Z^3	γ_3	-0.04659	-0.01884	-0.03144	-0.03869
	Z^4	γ_4	0.00947	0.00121	0.00967	-0.00079
Standard Deviation		σ	10.7128	10.4855	11.6032	10.9573

5.4.2.4 Pulse Rate in Relation to Age

Resting pulse rate will be compared to age-related norms [Fleming et al., 2011] according to the percentiles shown in Table 4.

Table 4 Percentiles for Resting Pulse Rate in Relation to Age and Gender

Age	Percentile						
	1st	10th	25th	50th	75th	90th	99th
4-6y	65	81	89	98	108	117	131
6-8y	59	74	82	91	101	111	123
8-12y	52	67	75	84	93	103	115
12-15y	47	62	69	78	87	96	108
15-18y	43	58	65	73	83	92	104

Source: web appendix to the Fleming et al article (web table 5)

Each value of pulse rate will be categorized as either being above the age-specific 99th percentile or below the 1st percentile.

5.4.2.5 Body Temperature

Body Temperature will be compared to a reference range of 35.4 to 37.7 degrees Celsius [Kliegman et al, 2007].

5.4.3 Clinical Laboratory Variables

Hematology, biochemistry and urinalysis assessments will be taken at visit 1/week -4 (screening), visit 5/week 4, and visit 7/week 12 (EoT). Additional hematology, biochemistry and urinalysis tests will be performed at visit 3/week 0 (baseline), and visit 8/week 14 (EoS) only if an AE related to hematology, biochemistry or urinalysis parameters occurred since the previous visit or at the discretion of the investigator.

All clinical laboratory assessments will be performed at a central laboratory or at a local laboratory if results can be made available to the investigative site (see protocol, Appendix 12.23, Table 11), except for urine pregnancy testing and urinalysis (urinalysis dipstick), which will be performed at a local laboratory. Pregnancy test in females of child bearing potential will be performed in serum at screening, at baseline, visit 5/week 4, visit 7/week 12 (EoT), and visit 8/week 14 (EoS).

Additional hematology, biochemistry and urinalysis (UA dipstick) tests will be performed at visit 3/week 0 (baseline), visit 5/week 4 and visit 8/week 14 (EoS) only if an AE related to hematology, biochemistry or urinalysis parameters occurred since the previous visit or at the discretion of the investigator. If any of the clinical laboratory tests results are outside the normal range at any scheduled time point during the study, the investigator may decide to repeat the test(s) on new samples. The clinical relevance of the abnormal results will be documented. Clinically relevant changes will be recorded as AEs (see Section 7.3 of protocol: Adverse Events and Other Safety Aspects).

The laboratory parameters that will be assessed during the conduct of the study are listed in Table 5.

Table 5 Laboratory Assessments

Hematology	Biochemistry	Urinalysis
Hematocrit Hemoglobin Mean corpuscular volume Mean corpuscular hemoglobin Mean corpuscular hemoglobin concentration Platelets Red blood cell count White blood cell count White blood cell count differential	Albumin Alanine aminotransferase Alkaline phosphatase Aspartate aminotransferase Bicarbonate Blood urea nitrogen Calcium Chloride Corrected serum calcium Creatinine Creatinine kinase Glucose Lactate dehydrogenase Magnesium Phosphate Potassium Serum hCG for female subjects Sodium Total bilirubin (total and direct) Total protein	Leukocyte esterase Nitrites pH Protein Red blood cells
hCG: human chorionic gonadotropin		

The value of each hematology and biochemistry parameter will be compared to its normal range and classified as High, Low or Normal. Urinalysis parameters will be similarly classified as either Normal or Abnormal.

When calculating changes or shifts in a result from baseline, the value from visit 1 will be used as baseline. If the visit 1 value is missing, the latest pre-baseline value will be used (i.e., from Visit 1 or any unscheduled visit conducted between visit 1 and 3).

5.4.4 Physical Examination

Physical examinations will be performed at visit 1/week -4 (screening), visit 5/week 4, and visit 7/week 12 (EoT) and will include assessments of the main body systems.

The physical examination will be performed per clinic standards and clinically significant findings at screening will be recorded as part of the subject's medical history. Clinically significant findings discovered after visit 1/screening will be recorded as an AE.

New or worsening clinically significant physical examination findings after signing of ICF will be recorded as AEs if they meet the criteria in Section 7.3 of protocol: Adverse Events and Other Safety Aspects.

5.4.5 12 Lead ECGs

A 12-lead ECG will be performed as single measurement at visit 1/screening, visit 3/baseline, visit 5/week 4, and visit 7/week 12 (EoT). ECG traces will be evaluated by the investigator who will give an overall interpretation and may leave a qualifying comment in the eCRF. The overall interpretation will be recorded as:

- Normal, or
- Abnormal - not clinically significant, or
- Abnormal - clinically significant

As well as the overall interpretation, the following ECG variables will be supplied by the central laboratory: PR Interval (msec), RR Interval (msec), QRS Duration (msec), QTcF interval (msec) and Heart Rate (HR) (beats/min).

5.4.6 Post Void Residual Volume

PVR volume will be assessed by ultrasonography at visit 3/week 0 (baseline), visit 5/week 4, visit 7/week 12 (EoT) and visit 8/week 14 (EoS). For each subject, the same method should be used throughout the study. The bladder should only be emptied when it was initially filled with preferably > 50% of the bladder capacity for age. Every attempt should be made to measure PVR within a minute of voiding.

A PVR of ≤ 20 mL (protocol version 1 and 2) is sufficient and the assessment does not have to be repeated. If the PVR is > 20 mL (Protocol version 1 and 2), the PVR assessment should be repeated (filled with preferably > 50% of the bladder capacity for age). If the subject is unable to complete a second measurement, it is up to the investigator to judge whether it can be skipped. At visit 3/week 0 (baseline), the lowest PVR result measured should be used to evaluate the exclusion criterion. In case the lowest PVR measured is > 20 mL (Protocol version 1 and 2) at visit 3/week 0 (baseline), the subject should be excluded from the study.

For subjects enrolled under protocol version 3.0 or higher (if applicable) the PVR volume should be 30 mL instead of 20 mL.

5.4.7 Acceptability and Palatability Questionnaires

The acceptability and palatability questionnaire will be completed at the end of treatment period. For each IP formulation, a separate questionnaire is available, see [Protocol Appendix 12.8 (Acceptability and Palatability Questionnaire for Tablets) and Appendix 12.9 (Acceptability and Palatability Questionnaire for Oral Suspension)].

5.4.7.1 Acceptability for Tablets

Subjects will evaluate the taste of the study medication ticking one of the following categories: “Really Bad” (0), “Bad” (1), “Not Bad, Not Good” (2), “Good” (3) and “Really Good” (4).

Subjects will evaluate the swallow of the study medication ticking one of the following categories: “Really Difficult” (0), “Difficult” (1), “Not Difficult, Not Easy” (2), “Easy” (3) and “Really Easy” (4).

Taste and swallow acceptability will be summarized as categorical variables.

5.4.7.2 Acceptability for Oral Suspension

Subjects will evaluate the taste of the study medication ticking one of the following categories: “Really Bad” (0), “Bad” (1), “Not Bad, Not Good” (2), “Good” (3) and “Really Good” (4).

Subjects will evaluate the smell of the study medication ticking one of the following categories: “Really Bad” (0), “Bad” (1), “Not Bad, Not Good” (2), “Good” (3) and “Really Good” (4).

Subjects will evaluate the consumption of the study medication ticking one of the following categories: “Really Difficult” (0), “Difficult” (1), “Not Difficult, Not Easy” (2), “Easy” (3) and “Really Easy” (4).

Subjects will evaluate the preparation of the study medication ticking one of the following categories: “Really Difficult” (0), “Difficult” (1), “Not Difficult, Not Easy” (2), “Easy” (3) and “Really Easy” (4).

Taste, smell, consumption and preparation acceptability will be summarized as categorical variables.

5.5 Other Endpoints

5.5.1 Exposure to Study Drug

The duration of exposure to each dose of study drug (PED25 and PED50) by visit and for the whole period will be calculated using the following information that is recorded in the eCRF:

- Overall start date and stop date of the study medication.

The first dose of study drug is to be administered on Day 1, the day after the baseline visit (visit 3). The initial dose will be PED25 for all subjects. The last dose of treatment is to be taken in the morning of visit 7/week 12/EoT.

- Start date and new dose of study drug at each titration step.

At visit 5/week 4 the dose may be up-titrated, or may remain the same.
Dose down-titration from PED50 to PED25 can be done at any time thereafter for safety reasons.

- Start date and new dose of study drug after each unscheduled dose change.
- At any time during the treatment period, the subject may have an unscheduled dose interruption or reduction. At each unscheduled change in dose, the (new) dose, start date and reason are recorded.

In addition, the return date of study medication is assumed to be the date of clinical visit.

In all cases where there is a dose change, either because of dose titration or an unscheduled dose interruption, reduction or increase, the last dose of the medication at the old level will be assumed to be the day before the given date of first dose at the new level. In this way, the subject's dosing history can be reduced to a series of unbroken intervals within each of which, the dose level is constant. From these, the subject's exposure at each dose can be calculated.

To illustrate, consider a subject with the following dose information on the eCRF:

- The dates of the very first and very last dose of the study medication given at D1=1st July 2019, Dlast=24th September 2019.
- Up-titration to PED50 at Visit 5. The first dose at the new level is given at D2=29th July 2019.
- Study drug interrupted at an unscheduled visit. The date of last dose at the PED50 level is given at D3=1st August 2019.
- Study drug re-started at PED25 at an unscheduled visit. The first dose at the new level is given at D4=5th August 2019.
- Remains at PED25 till Visit 6 (D5=25th August 2019). Up-titration to PED50 at Visit 6. First dose at the new level is given at D6=26thAugust 2019.

From these dates we calculate the following:

- The duration of exposure to the treatment can be calculated as ETOT=Dlast-D1+1-(D4-D3-1) =83 days.
- Between Visits 3 and 5, the subject is on PED25. The last dose at PED25 is given on D2-1, the day before the subject has the first dose at the new level PED50. The exposure at PED25 between Visits 3 and 5 is, therefore, (D2-1)-D1+1=28 days
- Between Visits 5 and 6, the study drug is temporarily suspended and then restarted. The subject will be considered to be exposed to the PED50 dose from

D2 until D3, i.e.
D3-D2+1=4 days.

- For the calculation of the exposure, suspension of the study drug is not ignored. The dose will be considered interrupted between D3+1 to D4-1, i.e. (D4-1) - (D3+1) + 1 = 3 days.
- The study drug is at PED25 from D4 to D6-1, the date of the last dose prior to Visit 6, i.e. (D6-1)-D4+1=21 days.
- Subject is on PED50 from D6 to Dlast, a total of Dlast-D6+1=30 days.
- The total exposure at PED50 is EPED50=34 days, at PED25 is EPED25=49 days
- The total exposure between Visits 3 and 5 is E1=28 days, between Visits 5 and 6 is E2=25 days, and between Visits 6 and 7 is E3 = 30 days.

5.5.2 Compliance to Study Drug

Tablet compliance

Compliance will be calculated according to the number of tablets of study medication dispensed, the number of wallets used and the duration of exposure between the applicable visits and overall.

Each wallet of study medication contains 35 tablets of 25 mg at visit 3/week 0, and 35 tablets of 50 mg and 7 tablets of 25 mg at visit 5/week 4. Subjects will receive an adequate number of wallets of both PED25 and PED50 as follows:

- At visit 3, 1 wallet containing 35 tablets of 25 mg will be dispensed
- At visit 5, wallets containing 35 tablets of 25 mg, or wallets containing 35 tablets of 50 mg and 7 tablets of 25 mg will be dispensed

The total number of tablets used between Visits i and j (i<j) is calculated as:

N_{used} = Total number of tablets dispensed at Visit i - Total number of tablets returned at Visit j

When a kit is not returned it will be assumed that all the medication from this kit was not used.

When a kit is returned late, it will be assumed that the medication was used during the study period that it was dispensed for, e.g. all the medication used from a kit dispensed at Visit 3 and returned at Visit 5 will be assumed to have been used between Visits 3 and 5.

The amount of expected study drug intake depends on the number of days of study drug treatment and the number of prescribed daily tablets. Since subjects are supposed to take 1 tablet per day, the total number of expected tablets to be used between Visits i and j is calculated as follows:

$N_{prescribed}$ = (Number of Exposure Days between Visit_i and Visit_j) x 1 tablet,

where Number of Exposure Days is calculated as: (Return date – Dispense date) + 1.

Between Visits i and j the compliance to study medication will be calculated as follows:

$$\text{Compliance} = (N_{used} / N_{prescribed}) \times 100\%.$$

To illustrate, at Visit 3 a subject was dispensed 1 wallet with 35 tables of 25 mg. The date of first dose was 1st July 2019. When the subject returned at Visit 5 (28th July 2019), the wallet contained 10 tablets.

$$N_{used} = (35\text{-}10) = 25 \text{ tablets}$$

$$\text{Number of exposure days} = (28\text{Jul}2019 - 01\text{Jul}2019) + 1\text{day} = 28 \text{ days},$$

$$\text{Hence, } N_{prescribed} = 28 \text{ tablets}$$

$$\text{Compliance between Visits 3 and 5} = (25/28) \times 100\% = 89.3\%.$$

For the whole study period, the total amount of study drug used ($TOTN_{used}$) is equal to the sum of the values of N_{used} at each applicable visit. Similarly, the total amount of study drug prescribed ($TOTN_{prescribed}$) is equal to the sum of the values of $N_{prescribed}$ at each applicable visit. Using these total values, the subject's compliance over the whole study period can be calculated as $100\% \times (TOTN_{used} / TOTN_{prescribed})$.

Suspension Compliance

Compliance for suspension will be calculated according to:

- Number of bottles dispensed at each visit (Table 6);
- Weight of the bottles returned to the site; and
- Duration of exposure between the applicable visits and overall.

Table 6 Number of Bottles to be Dispensed at each Visit

	Body Weight Range (kg)	Visit 3/Week 0	Visit 5/Week 4
PED25	13 to < 22	1	2‡
	22 to < 35	2	3‡
	≥ 35	2	4‡
PED50	13 to < 22	-	4
	22 to < 35	-	5
	≥ 35	-	7

PED25: pediatric equivalent dose 25 mg; PED50: pediatric equivalent dose 50 mg

Each kit of study medication contains a bottle with mirabegron granules. After filling the bottle containing the granules with 100 mL drinking water, the weight of the bottle with content will have a standard weight.

The total weight of the suspension/water and seal safes will not be weighed prior to dispensation but will be provided to the site and the IRT system will auto populate the weight of the IP prior to dispensation.

Each used bottle will be weighed upon return from the subject. Bottles that have been reconstituted and used will be weighed and the weight entered into the IRT drug accountability system. For returned bottles that are unused the initial starting weight will be utilized as a standard across all sites.

Doses are calculated weight-based, see Table 7. The body weight at Visit 3/baseline determines the weight range for the starting dose (PED25) and the up-titration dose (PED50) to be used.

Table 7 Body Weight-based Doses for Tablets or Suspension

	Body Weight Range† (kg)	Oral Suspension Volume‡ (mL)	Tablet Dose (mg)
PED25	13 to < 22	3	-
	22 to < 35	4	-
	≥ 35	6	25
PED50	13 to < 22	6	-
	22 to < 35	8	-
	≥ 35	11	50

PED25: pediatric equivalent dose 25 mg; PED50: pediatric equivalent dose 50 mg

† Subjects with a body weight ≥ 35 kg will receive the tablet. Subjects with a body weight< 35 kg or those who cannot be dosed with the tablet will receive an oral suspension.

‡ Oral suspension containing 8 mg/mL.

At each dispensing visit, subjects will receive an adequate number of kits of both PED25 and PED50 following the same procedure as descripted above for tablets.

The total dose of suspension used between Visits i and Visit j (i<j) is calculated as:

W_{used} = Total weight of bottles dispensed at Visit i - Total weight of bottles returned at Visit j, and taking into account the suspension strength of 8 mg/mL.

where:

- Total weight dispensed = Number of kits dispensed at the Visit i x bottle standard weight; and
- Total weight returned = Sum of the unused amount of study medication in each kit returned at Visit j.

When a kit is not returned it will be assumed that all the medication from this kit was not used.
When a kit is returned late, it will be assumed that the medication was used during the study period that it was dispensed for, e.g. all the medication used from a kit dispensed at Visit 3 and returned at Visit 5 will be assumed to have been used between Visits 3 and 5.

The amount of expected study drug intake depends on the number of days of study drug treatment and the number of prescribed daily suspension pouches. Since subjects are supposed to administer 1 dose per day, the total amount of expected suspension to be used between Visit i and Visit j is calculated as follows:

$W_{prescribed}$ = Duration in days between Visit i and Visit j times the subject's specific dose, see Table 7.

The weight of the subject's specific dose is determined by using a conversion factor of 1 mL corresponds to 1.0216 g.

The compliance to study medication between Visits i and j will be calculated as follows:

$$\text{Compliance} = (W_{used} / W_{prescribed}) \times 100\%.$$

To illustrate, a subject, who weighted 25 kg and with dose level PED25, was dispensed with 5 kits at Visit 3. The date of first dose was 1st July 2019. When the subject returned at Visit 4 (31th July 2019), 3 kits were not used and the weight of the 2 used kits was: 32 and 124 g. As an example, the bottle standard weight will be 137 g.

Compliance calculation:

1. Total amount of reconstituted study drug administered was resolved in (137-32) + (137-124) = 118 g of suspension, which is equal to 118/1.0216 mL = 115.5 mL.
2. Given that the solution strength is 8 mg/mL this means that in total 115.5 mL x 8 mg/mL = 924 mg of study drug was administered to the subject.
3. Duration was 31 days (= 31 July – 1 July +1), hence expected amount of study drug to be administered is 31 x 4 mL = 124 mL (based on the weight of the subject the suspension volume is 4 mL, see Table 7). This results in 124 mL x 8 mg/mL = 992 mg of study drug.
4. Compliance between Visits 3 and 4 = (924/992) x 100% = 93.15%.

For the whole study period, the total amount of study drug used ($TOTW_{used}$) is equal to the sum of the values of W_{used} at each applicable visit. Similarly, the total amount of study drug prescribed ($TOTW_{prescribed}$) is equal to the sum of the values of $W_{prescribed}$ at each applicable visit. Using these total values, the subject's compliance over the whole study period can be calculated as ($TOTW_{used}/TOTW_{prescribed}$) x 100%.

6 STATISTICAL METHODOLOGY

6.1 General Considerations

Efficacy data will be summarized for each age group (children [5 to <12 years of age] and adolescents [12 to <18 years of age]) by treatment group and visit, and safety data will be summarized for each age group and overall, by treatment group and visit, unless otherwise specified.

In general, continuous data will be summarized descriptively including the number of subjects (n), mean, standard deviation (SD), median, minimum and maximum. Categorical data will be summarized by frequencies and percentages. Percentages by categories will be based on the number of subjects with no missing data, i.e. the percentages for the non-missing categories will add up to 100%.

All statistical comparisons will be conducted using 2-sided tests at the 10% significance level unless stated otherwise. All null hypotheses will be of no treatment difference, all alternative hypotheses will be two-sided, unless specifically stated otherwise.

All data summarization and analyses will be performed using SAS® Version 9.4 or higher on Red Hat Enterprise Linux. Specifications for table, figure, and data listing formats can be found in the TLF specifications.

6.2 Study Population

6.2.1 Disposition of Subjects

The following subject data will be presented, for each of the two age categories, separately, and overall:

- Number of subjects with informed consent, discontinued before randomization, randomized (overall only) (All Screened Set)
- Number and percentage of subjects randomized in each analysis set, by treatment group and overall (All Randomized Set)
- Number and percentage of subjects completed and discontinued screening, by primary reason for treatment discontinuation (including COVID-19 reasons) for randomized subjects (overall only) (All Screened Set)
- Number and percentage of subjects completed and discontinued treatment, by primary reason for treatment discontinuation (including COVID-19 reasons) for randomized subjects, by treatment group and overall (All Randomized Set)
- Number and percentage of subjects completed and discontinued the follow-up period, by primary reason for post-study period discontinuation (including COVID-19 reasons) for randomized subjects and by treatment group and overall (All Randomized Set)
- Number and percentage of subjects excluded from PPS by reason for exclusion defined in Section 4.5.1, by treatment group and overall (FAS)
- Number and percentages of subjects per protocol version (All Screened Set).

A listing of inclusion and exclusion criteria, listing of first and last evaluations as well as a listing of subjects who were excluded from at least 1 analysis set will be provided for the All Randomized Set.

6.2.2 Protocol Deviations

The number and percentage of subjects with the following protocol deviation criteria will be summarized for each criterion and overall, for each age group and overall, by treatment group and total as well as by investigative site. Subjects deviating from a criterion more than once will be counted once for the corresponding criterion.

The unique identifiers will be as follows:

PD1 - Entered into the study even though they did not satisfy entry criteria

PD2 - Developed withdrawal criteria during the study and was not withdrawn

PD3 - Received wrong treatment or incorrect dose

PD4 - Received excluded concomitant treatment

PD5 - Clinical study procedure conducted prior to signing the initial consent form

PD6 - Events not reported within the expected turnaround time per regulatory reporting requirements

PD7 - Missed safety or efficacy assessments related to primary or key secondary endpoints

6.2.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized by descriptive statistics. Summaries will be presented, for each age group and overall, by treatment group.

Descriptive statistics for age, height (at baseline), weight (at baseline), BMI (at baseline, calculated as: weight (kg)/(height(cm)/100)^2), and EBC (=[age+1] × 30 in mL for children and 400 mL for adolescents) will be presented as well as frequency tabulations for sex, ethnicity, and race.
Note: for calculation of EBC the age of the subject in years [no decimals] will be used.
Summaries for height and weight with respect to age- and sex-specific percentiles will be provided. This will be done using SAS code and dataset stored on the Centers for Disease Control (CDC) and Prevention website:
https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm.
These summaries will be produced for the SAF, FAS and PPS.

Number and percentage of subjects randomized to treatment in each region, country and site will be presented, for each age group and overall, by treatment group for the SAF.
Geographical regions are defined as follows:

- Europe: Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Poland, Russian Federation, Serbia, Spain, Turkey, Ukraine, United Kingdom
- Asia and Pacific (APAC): Korea (Republic of), Malaysia, Philippines

- North America: United States, Canada and Mexico
- Southern Hemisphere: South Africa

The final list of participating/enrolling countries might be updated after end of recruitment when the list becomes final. If sites in additional countries will be opened due to changes in recruitment strategy, these countries will be allocated respectively and reported in the CSR.

Medical history is coded in MedDRA, and will be summarized by System Organ Class (SOC) and Preferred Term (PT) as well as by PT alone, for each age group and overall by treatment group for the SAF.

Demographic data and other baseline characteristics will be provided in listing format by country, site and subject for the All Screened Set.

6.2.4 Previous and Concomitant Medications

Previous and concomitant medications are coded with WHO-DD, and will be summarized by therapeutic subgroup (ATC 2nd level) and chemical subgroup (ATC 4th level) and preferred WHO name, for each age group and overall, by treatment group for the SAF.

Subjects taking the same medication multiple times will be counted once per medication and investigational period. A medication which can be classified into several chemical and/or therapeutic subgroups is presented in all chemical and therapeutic subgroups.

Previous medications are defined as medications that patients started and ended prior to first administration of study medication. Concomitant medications are defined as any medications that subjects took after the first dose of study medication and through Follow-up period (EoS). Medications that started prior to first administration of study drug and continued while study drug was given will be counted in both previous and concomitant medications.

Previous and concomitant medication data as well as non-medication therapy data will be provided in listing format by country, site and subject for the All Screened Set.

6.3 Study Drug Exposure and Compliance

6.3.1 Exposure

The following information on drug exposure will be presented, for each age group and overall category, by treatment group for the SAF:

- Descriptive statistics for cumulative amount of the drug to which the subject was exposed; and
- Number and percent of subjects with dose increases, decreases or interruptions.

Duration of exposure will be summarized in two ways.

- Descriptive statistics will be presented by age category and treatment group.
- Exposure time will be categorized according to the following categories by age category and treatment group:
  - less than 14 days
  - at least 14 days, less than 28 days
  - at least 28 days, less than 56 days

- 56 days or more
- Unknown.

Counts and percentages of subjects in each of these categories will be summarized, for each age group and overall, by treatment group for the SAF.

Exposure details will be provided in listing format by country, site and subject for the All Randomized Set.

6.3.2 Treatment Compliance

Overall compliance with the dosing schedule will be examined for subjects in the SAF whose total study drug count and first and last days of treatment are known.

Percent overall compliance will be summarized in two ways for the SAF:

- Descriptive statistics will be presented, for each age group and overall, by treatment group.
- Percent compliance will be categorized according to the following categories, for each age group and overall, by treatment group:
  - less than 70%
  - at least 70%, less than 80%
  - at least 80%, less than 120%
  - at least 120%, less than 130%
  - greater or equal than 130%
  - Unknown.

Treatment compliance details, including all data relevant to the calculation, will be listed by country, site and subject for the All Randomized Set.

6.4 Analysis of Efficacy

Efficacy analysis of the data from children (5 to <12 years of age) will be conducted on the FAS and PPS, and interpretation of results from statistical tests will be based on the FAS. The PPS will be used to assess the robustness of the results from the statistical tests based on the FAS.
Efficacy analysis in adolescents (12 to <18 years of age) will only be exploratory, and will be conducted on the FAS.

No multiplicity adjustment will be necessary in this study.

Efficacy data will be summarized for each age group by treatment and visit (week 2, 4, 8 and 12/EoT) unless otherwise specified. Efficacy data will be provided in listing format by country, site and subject for the All Randomized Set.

6.4.1 Analysis of Primary Efficacy Endpoint

The primary efficacy endpoint is change from baseline to week 12/EoT in mean micturitions per 24 hours in children (5 to <12 years of age).

6.4.1.1 Primary Analysis for Primary Efficacy Endpoint

The primary estimator for the primary estimand as defined in Section [2.1.1 Estimand] will be calculated according to the evaluation of the primary efficacy endpoint. The primary efficacy endpoint will be analyzed using an analysis of covariance (ANCOVA). The ANCOVA model will include treatment group, sex and geographical region as fixed effects and the mean number of micturitions per 24 hours at baseline as covariate. This analysis will be performed with imputation of missing visit 7/week 12 data using the last observation carried forward (LOCF) method.

The hypothesis for comparisons is given as follows:

H_0: The adjusted mean change from baseline at visit 7/week 12 in mean number of micturitions per 24 hours for mirabegron and placebo are the same.

H_1: The adjusted mean change from baseline at visit 7/week 12 in mean number of micturitions per 24 hours for mirabegron and placebo are not the same.

Statistical testing will be done 2-sided at a 10% significance level. A 90% CI for the treatment difference (mirabegron – placebo) will be provided. The primary analysis will use the FAS. The plausibility of the underlying assumption of missing at random for the primary estimator will be evaluated and, therefore, the estimation of the treatment effect.

The ANCOVA model results will be presented with least squares (LS) means and two-sided 90% confidence intervals (CIs) for mean changes from baseline within each treatment group. Differences in LS means between mirabegron versus placebo will be derived and presented together with two-sided 90% CIs. T-statistics corresponding to the Type III sums of squares for the differences in the LS means will be used to obtain p-values for the comparisons.

Model fitting will be visually assessed. A scatter plot of residuals versus predicted values, along with histogram and normal probability plots will be produced.

Sensitivity analyses under alternative assumptions for missing data mechanisms such as missing not at random to show robustness of the estimates will be further specified.

6.4.1.2 Graphical Summary of Data from Primary Efficacy Endpoint

Mean number of micturitions per 24 hours at each visit and change from baseline in mean number of micturitions per 24 hours (without LOCF) will be plotted for FAS: mean (90% CI) values will be presented for each treatment group and visit.

6.4.1.3 Sensitivity Analysis for Primary Efficacy Endpoint

The following sensitivity analyses of the primary efficacy endpoint will be produced.

6.4.1.3.1 Analysis with LOCF for PPS

The same analysis as described in Section 6.4.1.1 will be repeated for PPS (with imputation of missing data, LOCF).

6.4.1.3.2 Analysis without LOCF

The same analysis as described in Section 6.4.1.1 will be repeated for both FAS and PPS, both without imputing for missing data.

6.4.1.3.3 Analysis using Repeated Measures ANCOVA

The primary endpoint will also be analyzed using a repeated measurements analysis of covariance (ANCOVA) model with treatment group, visit (week 2, 4, 8 and 12), sex, geographical region and the interaction between treatment group and visit as fixed effects and mean number of micturitions per 24 hours at baseline as covariate.

An appropriate covariance structure will be selected for the within subject repeated measurements. An unstructured correlation pattern will be used to estimate the variance-covariance of the within-subject repeated measures at first instance. Parameters will be estimated using Restricted Maximum Likelihood with the Newton-Raphson algorithm and using the Kenward-Roger method for calculating the denominator degrees of freedom. If there is a convergence problem due to the unstructured covariance matrix, the unstructured covariance matrix will be replaced by the compound symmetry covariance matrix.

Mirabegron will be compared to placebo using a linear contrast within the repeated measurements ANCOVA model, with 2-sided significance level of alpha = 0.1 and 90% CI.

This analysis will be conducted on the FAS and will serve as a sensitivity analysis to the LOCF method used in the primary model to assess the robustness of the findings.

Model fitting will be visually assessed. A scatter plot of residuals versus predicted values, along with histogram and normal probability plots will be produced.

6.4.1.3.4 Nonparametric Analysis

A nonparametric ANCOVA on rank-transformed data (changes from baseline) will be performed (stratified rank ANCOVA) (with LOCF). Statistical hypothesis testing (p-values) will be performed: the p-value of the treatment difference will be assessed using a rank ANCOVA stratified by region.

Ranks within each geographical region will be derived across the two treatment groups for the baseline value and the change from baseline value.

Then separate linear regression models for each geographical region will be performed with response variable ranked change from baseline value and dependent variables sex and ranked baseline value.

Finally, the stratified mean value test, using the values of the residuals from the regression analysis as scores and geographical region and sex as stratum, compares the two treatment groups using the Cochran-Mantel-Haenszel (CMH) statistics.

Appendix 9.1 present an example of the SAS® code to perform this analysis.

In addition, a (nonparametric) estimate for the difference between placebo and mirabegron, with associated (asymptotic) 90% CI, according to Hodges-Lehmann and Moses, will be presented.

These analyses will be conducted on the FAS.

6.4.1.3.5 Sensitivity Analysis for the Primary Estimand

Two analyses of the primary endpoint will be performed where missing data will be imputed using multiple imputation obtained from the procedure MI from SAS®.

The imputation process will be broken up into a sequence of multiple imputation steps using PROC MI. Two different sensitivity analyses will be performed by following either steps 1, 2a, 3, and 4 (missing not at random), or steps 1, 2b, 3, and 4 (missing at random):

Step 1: Subjects who have missing values for intermediate visits, the missing data will be imputed using the Monte Carlo Markov Chain (MCMC) methodology, as available in PROC MI and option MCMC. The results of this procedure will be stored in a dataset. This dataset now contains a so-called monotone missing data pattern (i.e., for a subject with missing data, all values are missing after a certain time-point). It is assumed here that subjects with missing data follow the same model as other subjects in their respective treatment arm that have complete data. It can be said that this is a reasonable assumption for this partial imputation process because subjects tend to miss intermediate visits due to scheduling conflicts or other reasons unrelated to their medical condition under study.

Step 2a: sequentially for one time-point data will be imputed for missing values: first for week 2 -> week 4 -> week 8 and finally for week 12.
Missing data will be multiple imputed under a missing not at random assumption: borrow information from placebo arm subjects. When imputing missing values for week t (t = 2, 4, 8, 12), the input dataset will include data from all placebo subjects and only data from those subjects from the mirabegron arm that have missing values at week t (Go to step 3.).

Step 2b: sequentially for one time-point data will be imputed for missing values: first for week 2 -> week 4 -> week 8 and finally for week 12.
Missing data will be multiple imputed under a missing at random assumption: borrow information from patients in the same treatment arm. When imputing missing values for week t (t = 2, 4, 8, 12), the input dataset will include data from all subjects and only data from those subjects from the placebo and mirabegron arm that have missing values at week t.

Step 3: Perform ANCOVA on each of the imputed datasets at week 12.

Step 4: Results from ANCOVA analysis on the imputed datasets will be combined to derive an overall result by using PROC MIANALYZE.

These analyses will be conducted for the FAS.

6.4.1.4 Summary by Subject and Caregiver

In this study the eDiary contains an entry to capture whether the caregiver or the subject entered the data.
To investigate whether there is a difference in scoring the micturitions by the caregiver or the subject the mean number of micturitions per 24 hours will be calculated for each of them separately. This will be done as follows: for each (valid diary) day, the number of micturitions entered in the eDiary will be calculated for the subject and the caregiver, separately. Next, for each visit, the mean number of micturitions will be calculated for the caregiver and subject (see Section 5.1).
Presented will be the results for each age group and overall by treatment, visit and caregiver/subject.
Note 1: a diary for a caregiver is valid for an analysis visit when at least 2 valid diary days are available. Same for subject.
Note 2: in the event that 2 or more caregivers enter data in the eDiary of a subject these data will be reported as coming from 1 caregiver.

6.4.1.5 Subgroup Analysis for Primary Efficacy Endpoint

For the primary efficacy endpoint, descriptive statistics (n, mean, 95% confidence interval around the mean, SD, minimum, median, maximum) will be calculated for mean change from baseline at each visit by sex, race, ethnicity, geographical region, formulation (tablets vs. oral suspension), prior OAB treatment and symptomatic UTI, using LOCF.

The following interactions will be investigated, separately:

- treatment group by sex;
- treatment group by geographical region;
- treatment group by formulation;
- treatment group by prior OAB treatment; and
- treatment group by symptomatic UTI (yes/no)
  Note: symptomatic UTI is yes when it is reported as an AE during the trial, irrespective when it occurs.

To investigate the first two interactions, separately, data of the primary efficacy endpoint will be analyzed using an ANCOVA model with treatment group, sex and geographical region as fixed effects, the mean number of micturitions per 24 hours at baseline as covariate and including the, above mentioned, interaction term that is investigated.

To investigate the last three interactions, separately, data of the primary efficacy endpoint will be analyzed using an ANCOVA model with treatment group, sex, geographical region and either formulation, prior OAB treatment, or symptomatic UTI (yes/no) as fixed effects, the mean number of micturitions per 24 hours at baseline as covariate and including the, above mentioned, interaction term that is investigated.

6.4.2 Analysis of Secondary Efficacy Endpoints

Descriptive statistics (n, mean, SD, minimum, median, maximum) and the same primary efficacy analysis (ANCOVA) as for the primary endpoint will be applied to change from baseline at the end of the 12-week treatment period in children (5 to <12 years of age) for:

- Mean volume voided per 24 hours
- Maximum volume voided
- Mean number of daytime incontinence episodes per 24 hours
- Mean number of daytime micturitions per 24 hours

Change from baseline in the number of dry (incontinence free) days per 7 days at the end of the 12-week treatment period will be analyzed with a negative binomial regression model including treatment group, sex, geographical region as factors and the log baseline rate of number of dry days (e.g., the log of the ratio of dry days and number of diary days at baseline) as covariate.

Data of these secondary efficacy endpoints will be analyzed both without imputation and with imputation using the LOCF method, for the FAS.

The same analysis detailed in Section 6.4.1.3 will be performed for all the secondary efficacy endpoints on the FAS as sensitivity analysis.

6.4.3 Analysis of Exploratory Endpoints

Data of the following explorative endpoints will be summarized, separately, for pediatric subjects:

- Percentage of subjects with a reduction in daytime incontinence episodes (< 50% reduction [non-responder], 50% [partial responder] and 100% [responder])
- Improvement from baseline in worst incontinence grading. Average over 7 days will be calculated of the score of "What was your worst leakage during the day": 1 - Slightly Wet; 2 - Moderately Wet; and 3 - Fully Wet
- Mean number of micturitions per 24 hours, adjusted for fluid intake:
  - Number of micturitions per 24 hours that would be required at week 12/EoT to void the same total volume as voided at baseline
  - Difference in mean number of micturitions per 24 hours adjusted for fluid intake compared to the expected mean number
- Mean volume voided per micturition as a percentage of expected bladder capacity
- Categorized mean volume voided per micturition as a percentage of expected bladder capacity (results and changes from baseline).

Efficacy endpoints for adolescents will be analyzed descriptively.

- Change from baseline at the end of the 12-week treatment period (adolescents only):
  - Mean number of micturitions per 24 hours
  - Mean number of daytime micturitions per 24 hours
  - Mean volume voided per 24 hours

- Maximum volume voided
- Mean number of incontinence episodes per 24 hours

- Number of dry (incontinence-free) days per 7 days at the end of the 12-week treatment period (adolescents only)
- Mean number of daytime grade 3 or 4 on Patient Perception of Intensity of Urgency Scale (PPIUS) urgency episodes per 24 hours (adolescents only) will be summarized by treatment group and visit.

Correlations between the primary endpoint and mean volume voided, as well as incontinence episodes, will be explored. Scatter plots will be presented to access the correlation between the primary end point and mean volume voided and number of incontinence episodes at visit 7/week 12.

6.5 Analysis of Safety

Safety analysis will be conducted on the SAF and conducted separately for children, adolescents and overall. Safety endpoints will be summarized, for each age group and overall, by treatment group and visit, if applicable. Presented will be descriptive statistics.

6.5.1 Adverse Events

Summaries and listings of SAEs and Serious TEAEs include SAEs upgraded by the sponsor based on review of the Sponsor's list of Always Serious terms or upgraded according to the Important Medical Event process, if any upgrade was done.

The coding dictionary for this study will be MedDRA. It will be used to summarize AEs by SOC and PT.

An overview table will include the following details:

- Number of TEAEs
- Number and percentage of subjects with TEAEs
- Number of drug related TEAEs
- Number and percentage of subjects with drug related TEAEs
- Number of serious TEAEs
- Number and percentage of subjects with serious TEAEs
- Number of serious drug related TEAEs
- Number and percentage of subjects with serious drug related TEAEs
- Number of TEAEs leading to permanent discontinuation of study drug
- Number and percentage of subjects with TEAEs leading to permanent discontinuation of study drug
- Number of drug-related TEAEs leading to permanent discontinuation of study drug
- Number and percentage of subjects with drug-related TEAEs leading to permanent discontinuation of study drug
- Number of TEAEs leading to death
- Number and percentage of subjects with TEAEs leading to death
- Number of drug-related TEAEs leading the death

- Number and percentage of subjects with TEAEs leading to death
- Number and percentage of subjects who died.

The number and percentage of subjects with TEAEs, as classified by SOC and PT will be summarized by age group and overall for each treatment group. Summaries will be provided for:

- TEAEs
- Drug related TEAEs
- Serious TEAEs
- Drug related serious TEAEs
- TEAEs leading to permanent discontinuation of study drug
- Drug related TEAEs leading to permanent discontinuation of study drug
- Common (>=5% in Any Treatment Group) TEAEs Excluding SAEs

The number and percentage of subjects with TEAEs, as classified by PT only, will be summarized by age group and overall for each treatment group.

The number of TEAEs and the number and percentage of subjects with TEAEs, as classified by SOC and PT will also be summarized by severity and by relationship to study drug. In the subject count, if a subject has multiple TEAEs with the same SOC or PT, but with differing severity or relationship, then the subject will be counted only once with the worst severity and highest degree of relationship, however, if any of the severity or relationship values are missing then the subject will be counted only once with missing severity or relationship. In the adverse event count, the adverse events will be presented in each category they were classified to. Drug related TEAEs will be presented in a similar way by severity only.

Adverse events of interest will be tabulated by PT.

AEs will be provided in listing format by country, site and subject for the SAF.

6.5.2 Clinical Laboratory Evaluation

The baseline value will be the last non-missing value taken on or prior to first dose of study drug.

Quantitative clinical laboratory variables, i.e., hematology, biochemistry, and urinalysis will be summarized using mean, standard deviation, median, minimum and maximum, for each age group and overall, by treatment group and visit.
Summary of laboratory values will be presented for visit 1, 3, 5, 7 and 8 (EoS)], and for EoT values.

Additionally, a within-subject change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way. Each laboratory result will be classified as low (L), normal (N), or high (H) at each visit according to the laboratory supplied reference ranges.

The number and percentage of subjects below and above reference range will be summarized, for each age group and overall, by treatment group and visit.

Frequency tabulations of qualitative clinical laboratory variables (urinalysis) will be presented, for each age group and overall, by treatment group and visit.

Shifts from baseline to week 4, week 12 and EoT will be summarized by two types of shift tables:

- Shift tables of reference range (low, normal, high) changes from baseline to week 12 and to EoT, and
- Summary shifts of reference range changes from baseline to week 4, week12 and EoT.

These shifts are categorized as:

- "Shift to Low": shift from normal or high to low
- "Shift to High": shift from normal or low to high
- "Categorized Increase": shift from low to normal or from normal to high
- "Categorized No Change": value stays in the same reference range
- "Categorized Decrease": shift from high to normal or from normal to low.

All clinical laboratory data collected during the study and variables derived from it will be listed using SAF.

6.5.2.1 Central and Local Laboratory Data

In this study laboratory data may come from a central laboratory or a local laboratory, see Table 12 of the protocol.

For summarizing the data of a laboratory parameter the data from the local laboratory will be transformed. First all data from the parameter assessed in the local laboratory will be transformed into SI units. Next local laboratory data will be transformed according to the transformation formula below. Let x be the value of the laboratory parameter from the local laboratory, ULNL the upper limit of the normal range of the local laboratory, and ULNC the upper limit of the normal range of central laboratory then the transformed value (t) from the local laboratory is:

$$t = x.\frac{ULNC}{ULNL}$$ (Scale normalization, Karvanen et al.)

This value t will be used for reporting laboratory values.

The results from the local laboratory will be used to determine whether a result is outside the normal range.

6.5.2.2 Liver Safety Assessment

The liver safety assessments will be summarized by the categories below based on the measurements from Alkaline Phosphatase (ALP), Alanine Transaminase (ALT), total bilirubin, Aspartate Transaminase (AST) and their combination. These parameters will be based on measurements from a central laboratory.

The subject's highest value during the investigational period will be used.

- ALT >3xULN, >5xULN, >10xULN, >20xULN
- AST >3xULN, >5xULN, >10xULN, >20xULN
- ALT or AST >3xULN, >5xULN, >10xULN, >20xULN
- ALP >1.5xULN
- Total Bilirubin >2xULN
- (ALT or AST >3xULN) and Total Bilirubin >2xULN
- (ALT or AST >3xULN) and ALP <2xULN and Total Bilirubin >2xULN

The last 2 criteria where 2 or more parameters are evaluated will be with the measurements on the same day or up to 1 day apart (Samples to be taken simultaneously or within a maximum of 24 hours).

The denominator for each criterion will be the number of subjects who have at least one value during the investigational period. The number and percentage of subjects meeting the criteria during the investigational period will be summarized, within and across age group by treatment group.

A scatter plot of Peak AST or ALT data vs. Peak Total Bilirubin data will be presented for mirabegron subjects. Data will be presented as fractions of the ULN. In this scatter plot the area for Hy's Law will be presented, where ALT or AST >3xULN and Total bilirubin > 2xULN (ALT, ALT and Total Bilirubin from same blood sample or within a maximum of 24 hours). The area for hyperbilirubinemia (ALT and AST <3xULN and Total bilirubin >2xULN), Temple's corollary (ALT or AST >3xULN and Total bilirubin < 2xULN) will also be presented in this scatter plot.

Laboratory data will be provided in listing format by country, site and subject for the SAF.

6.5.3 Vital Signs

The baseline value will be the last non-missing value taken on or prior to first dose of study drug.

In addition to the presentation by age groups 5 to < 12 years of age, 12 to < 18 years of age and overall, vital signs will also be presented for age groups 5 to < 8 years of age and 8 to < 12 years of age.

6.5.3.1 Clinical Measurements

Values and changes from baseline for vital signs (SBP, DBP, pulse rate and measurement for temperature) and percentiles of SBP, DBP and Pulse Rate compared to age and height norms (see Appendix 9.3 for Blood Pressures and Table 4 for Pulse Rate) will be listed and summarized, for each age group and overall, by treatment group and visit using mean, standard deviation, median, minimum and maximum. Summary of vital signs will be presented for each visit [visit 1, 3, 5, 7 and 8 (EoS)], and for EoT values.

2017 American Academy of Pediatric Clinical Practice Guidelines

For each scheduled post-baseline visit, shift tables from baseline to each post-baseline visit will be presented with respect to the changes from normal blood pressure to elevated blood pressure; stage 1 HTN, or stage 2 HTN, see Table 8. For the cut-off points (90th, and 95th percentiles) see 2017 American Academy of Pediatric Clinical Practice Guidelines.

Table 8 Definitions of Blood Pressure Categories and Stages

Children Aged 1–<13 years	Adolescents Aged ≥13 years
Normal BP: <90th percentile	Normal BP: <120/<80 mm Hg (SBP < 120 mmHg and DBP < 80 mmHg)
Elevated BP: ≥90th percentile to <95th percentile or 120/80 mmHg to <95th percentile (whichever is lower)	Elevated BP: 120/<80 to 129/<80 mmHg (120 ≤ SBP ≤ 129 mmHg and DBP ≤ 80 mmHg)
Stage 1 HTN: ≥95th percentile to <95th percentile + 12 mmHg, or 130/80 to 139/89 mmHg (whichever is lower)	Stage 1 HTN: 130/80 to 139/89 mmHg (130 ≤ SBP ≤ 139 mmHg or 80 ≤ DBP ≤ 89 mmHg)
Stage 2 HTN: ≥95th percentile + 12 mmHg, or ≥140/90 mmHg (whichever is lower)	Stage 2 HTN: ≥140/90 mmHg (SBP ≥ 140 or DBP ≥ 90 mmHg)

Table obtained from: 2017 American Academy of Pediatric Clinical Practice Guidelines, Table 3
Note: the age classification into children and adolescents in the Clinical Practice Guidelines (2017) differs from the definition used in the protocol

Summaries will be presented for SBP and DBP separately, for each age group and overall, by treatment group. See further details in Section 5.4.2.3: Conversion of Blood Pressure to Percentiles.

In addition, PR measurements will be compared to age-related norms [Fleming et al, 2011] according to the percentiles shown in Table 4. Each PR value will be categorized as either below the 1st percentile, between the 1st and 99th percentile (limits included) or above the 99th percentile. For each scheduled post-baseline visit, shift tables from baseline to each post-baseline visit will be presented with respect to the changes from PR to below the 1st percentile and above the 99th percentile.

Summaries will be presented, for each age group and overall, by treatment group for the number of subjects with PR values below and above the normal range, the normal range is defined as: 1st – 99th percentile.

Potentially Clinically Relevant (PCR) criteria

PCR criteria for SBP will be identified using the following criteria and cut-points:

- Hypertension SMQ [20000147] (narrow search), or
- 2017 American Academy of Pediatric Clinical Practice Guidelines:
  - For subjects aged 5 - <13 years: any SBP above the 95th percentile +12 mmHg or BP≥ 140/90 mmHg at a post baseline visit, and
    for subjects aged ≥13 years: BP ≥140/90 mmHg at a post-baseline visit,
    or

- Change in category from baseline (Normal or elevated BP or Stage 1 HTN) to post-baseline (Stage 1 HTN or Stage 2 HTN).

PCR criteria for DBP will be identified using the following criteria and cut-points:

- Hypertension SMQ [20000147] (narrow search), or
- 2017 American Academy of Pediatric Clinical Practice Guidelines:
  - For subjects aged 5 - <13 years: any DBP above the 95th percentile + 12 mmHg or BP≥ 140/90 mmHg at a post baseline visit, and
    for subjects aged ≥13 years: BP ≥140/90 mmHg at a post-baseline visit,
    or
  - Change in category from baseline (Normal or elevated BP or Stage 1 HTN) to post-baseline (Stage 1 HTN or Stage 2 HTN).

Using these criteria, a high and low flag will be created to indicate if values meet these conditions (i.e., SBP/DBP PCR values) or not (i.e., non-SBP/DBP PCR values).

For PR, the below criterion will be used to identify a PCR increase:

- PR above 99th percentile compared to age-related norms PR [Fleming et al, 2011] and ≥15 bpm change from baseline.

PCR for SBP, DBP and PR will be summarized, for each age group and overall, by treatment group.

Data obtained from the clinical measurements (SBP. DBP and PR) will be provided in listing format by country, site and subject for the SAF.

6.5.3.2 Self-Measurements

Similarly to the clinic vital signs measurements (Section 6.5.3.1), values and changes from baseline for self vital signs measurements (SBP, DBP, pulse rate) and percentiles of self-measurements SBP and DBP compared to age and gender norms will be listed and summarized at each visit using mean, standard deviation, median, minimum and maximum. The number and percentage of subjects outside reference ranges will be shown.

For the purpose of the analyses, the average self-measurement vital signs will be calculated using all values from the 2-days diary collection for the regular measurements. If all measurements are missing, then the average will be missing. Extra measurements will be listed only.

For shifts from baseline to each post-baseline visit the categories used for presentation were obtained from the 2017 (American Academy of Pediatric) Clinical Practice Guidelines. For each scheduled post-baseline visit, shift tables from baseline to each post-baseline visit will be presented. Categories are: normal blood pressure; elevated blood pressure; stage 1 HTN, and stage 2 HTN, see Table 8. For the cut-off points (90th, and 95th percentiles) see 2017 American Academy of Pediatric Clinical Practice Guidelines.

Potentially Clinically Relevant (PCR) criteria

PCR criteria for SBP will be identified using the following criteria and cut-points:

- Hypertension SMQ [20000147] (narrow search), or
- 2017 Clinical Practice Guidelines:
  - For subjects aged 5 - <13 years: Any SBP above the 95th percentile +12 mmHg or BP≥ 140/90 mmHg at a post baseline visit, and
    for subjects aged ≥13 years: BP ≥140/90 mmHg at a post-baseline visit,
    or
  - Change in category from baseline (Normal or elevated BP or Stage 1 HTN) to post-baseline (Stage 1 HTN or Stage 2 HTN).

PCR criteria for DBP will be identified using the following criteria and cut-points:

- Hypertension SMQ [20000147] (narrow search), or
- 2017 Clinical Practice Guidelines:
  - For subjects aged 5 - <13 years: Any DBP above the 95th percentile +12 mmHg or BP≥ 140/90 mmHg at a post baseline visit, and
    for subjects aged ≥13 years: BP ≥140/90 mmHg at a post-baseline visit,
    or
  - Change in category from baseline (Normal or elevated BP or Stage 1 HTN) to post-baseline (Stage 1 HTN or Stage 2 HTN).

Using these criteria, a high and low flag will be created to indicate if values meet these conditions (i.e., SBP/DBP PCR values) or not (i.e., non-SBP/DBP PCR values).

PR measurements will be compared to age-related norms [Fleming et al, 2011] according to the percentiles shown in Table 4. For each scheduled post-baseline visit, shift tables from baseline to each post-baseline visit will be presented with respect to the changes from PR to below the 1st percentile and above the 99th percentile.

For PR, the below criterion will be used to identify a PCR increase:

- PR above 99th percentile compared to age-related norms PR [Fleming et al, 2011] and ≥15 bpm change from baseline.

Tables for potentially clinically relevant vital signs will be produced using the same criteria described in Section 6.5.3.1.

Data obtained from the self-measurements (SBP. DBP and PR) will be provided in listing format by country, site and subject for the SAF.

6.5.4 Electrocardiograms

ECG variables will be summarized using mean, standard deviation, median, minimum, and maximum, for each age group and overall, by treatment group and visit, including changes from baseline. Summary of ECG data will be presented for visits 1, 3, 5, 7, and for EoT values.

Number and percent of subjects with normal, not clinically significant abnormal, and clinically significant abnormal results as assessed by investigator for the 12 lead ECG will be tabulated, for each age group and overall, by treatment group and visit.

Shift tables will be presented for changes from baseline at each visit including local interpretation categories. Shift tables will also be presented as above from baseline to worst post-baseline value during the treatment period.

The corrected QT interval (i.e. QTcF) will be summarized for each visit showing the number and percentage of subjects in each of the following categories:

For the children category:

- 440 ms

For the male adolescents category:

- 450 ms
- 500 ms

For the female adolescents category:

- 480 ms
- 500 ms

Note that these categories are cumulative in that subjects satisfying criterion for more extreme category will also be counted in each applicable less extreme category.

The corrected QT interval (i.e., QTcF) will also be summarized, for each age group and overall, by treatment group and visit showing the frequencies of subjects with the following changes from baseline: <0, ≥0 and <30, ≥30 and <60, and ≥ 60 ms.

Number and percent of subjects with 12 lead ECG abnormalities as well as number and percent of subjects whose 12 lead ECG reading changed from normal at baseline to abnormal will be tabulated, for each age group and overall, by treatment group and visit.

ECG data will be provided in listing format by country, site and subject for the SAF.

6.5.5 Pregnancies

A detailed listing of all pregnancies will be provided if any occur using All Screened Set.

6.5.6 Other Safety-Related Assessments

6.5.6.1 Post Void Residual Volume

PVR will be summarized, for each age group and overall, by treatment group and visit using mean, standard deviation, median, minimum, and maximum at each treatment visit, including changes from baseline. In case of 2 measurements at the same visit, the lowest (instead of mean) PVR volume will be used for the calculations.
Summary of PVR data will be presented for visits 3, 5, 7 and 8 (EoS), and for EoT values.

To calibrate PVR with bladder size, the ratio of PVR to the child's Expected Bladder Capacity (EBC, =[age+1] × 30 in mL for children and 400 mL for adolescents) will also be

calculated at baseline and post-baseline study visits and expressed as a percentage. These ratios will be summarized together with the change from baseline by age group and overall.

The following variables will be summarized, for each age group and overall, by treatment group at each visit (and also for the maximum recorded post-baseline value of PVR) using number and percentage of subjects in the following categories:

- PVR: 0 to <20 mL, 20 to <50 mL, ≥50 mL
- Change from baseline in PVR: <0 mL, 0 to <20 mL, 20 to <50 mL, ≥50 mL
- Ratio of PVR to EBC (%): 0% to <20%, 20% to <40%, ≥40%
- Change from baseline in ratio of PVR to EBC (%): <0%, 0% to <20%, 20% to <40%, ≥40%

PVR data will be provided in listing format by country, site and subject for the SAF.

6.5.6.2 Body Weight and Height

Body weight, height and BMI will be summarized, for each age group and overall, by treatment group and visit using mean, standard deviation, median, minimum, and maximum at each visit, including changes from baseline. Summary of body weight, height and BMI will be presented for visits 1, 3 and 7, and for EoT values.
Body weight, height and BMI data will be provided in listing format by country, site and subject for the SAF.

6.5.6.3 Acceptability and Palatability Questionnaires

Results from the acceptability and palatability questionnaire for tablets (see protocol Section 12.8) and for oral suspension (see protocol Section 12.9) will be summarized at week 12/EoT, for each age group and overall, by treatment group.

Data obtained from the questionnaires will be provided in listing format by country, site and subject for the SAF.

6.5.7 Subgroups of Interest for Safety

In addition to the analyses described above, selected safety variables (TEAEs, drug-related TEAEs and Vital Signs) will be summarized for each age group and overall by treatment group for several subgroups.

For TEAEs and drug-related TEAEs the subgroups are: sex, race, ethnicity, geographical region, formulation (tablets vs. oral suspension), prior OAB treatment and symptomatic UTI, and for vital signs: sex, race, ethnicity, and formulation (tablets vs. oral suspension).
For any of the subgroups specified for vital signs, at least 10 subjects by stratum are required.

6.6 Analysis of Pharmacokinetics

Plasma concentrations for mirabegron will be sampled at visit 5/week 4 and visit 7/week 12 (EoT) (Table 9). One predose sample will be collected on both pharmacokinetic sampling days. On pharmacokinetic sampling days, dosing should occur in the clinic and breakfast should be eaten at the clinic within 1 hour before dosing.

Table 9 Sample Collection Schedule

Visit	Time Point	Collection Window	Whole Blood Pharmacokinetics
Visit 5/week 4	Predose (trough)	Approximately 24 hours after previous dose	X
Visit 7/week 12 (EoT)	Predose (trough)	Approximately 24 hours after previous dose	X

EoT: end of treatment

Pharmacokinetic analysis will be conducted on the PKAS and conducted separately for children, adolescents and overall. Descriptive statistics (n, mean, SD, minimum, median, maximum, coefficient of variation (CV), geometric mean and geometric CV) will be used to summarize plasma concentrations of mirabegron by visit. Plasma concentration values below the lower limit of quantification (LLOQ, 0.0500 ng/mL) will be set equal to 0 in the calculation of summary statistics.

The following plot will be produced: Box plots of plasma concentration-time profiles (normal scale and semi-log scale).

Further details will be described in the TLF specifications.

Pharmacokinetic analysis will be conducted on the PKAS and parameters will be summarized for children, adolescents, and overall in the modeling report. Results of the pharmacokinetic analysis will be provided in a separate PK modeling report.

6.6.1 Statistical Analysis

Plasma concentration data from this study will be pooled with other pediatric PK data and will be analyzed using a population modeling approach. Details of the pharmacokinetic analysis will be presented in a separate population PK modeling analysis plan.

6.7 Analysis of Pharmacodynamics

Not applicable.

6.8 Other Analyses

Not applicable.

6.9 Interim Analysis (and Early Discontinuation of the Clinical Study)

One comparative interim analysis will be performed after 50% of children planned to be randomized have had their week 12/EoT assessment. Only data from those children with week 12/EoT assessment will be used in the interim analyses. The interim analysis will determine if the chance of a positive study with respect to the primary endpoint at the final analysis is high enough to justify continuation of the study; otherwise, the study will be stopped for futility.

Given the data at the interim analysis, a predictive probability of a positive study will be calculated. A positive study is defined as the final 1-sided p-value ≤ 0.05 when comparing mirabegron vs. placebo in the primary efficacy analysis (in children [5 to < 12 years of age]).

If the predictive probability is ≤ 5% at this interim evaluation, the study will be recommended to stop for futility. Otherwise, the study will continue.

No efficacy analyses will be performed except for determination of futility.

At the interim analysis, the following predictive probability will be calculated:
Predictive Probability of ({final one-sided p-value of mirabegron vs. placebo ≤ 0.05} |Data).

Since we use non-informative prior distribution of difference in mean change from Baseline to week 12/EoT in mean number of micturitions per 24 hours between the mirabegron group and the placebo group and assume the mean change from Baseline to week 12/EoT in mean number of micturitions per 24 hours follows a Gaussian distribution, the Bayesian predictive probability for the event of a final one-sided p-value of mirabegron vs. placebo ≤ 0.05 (denoted by *PP*) can be obtained using a closed form based on the methods described in Geisser and Johnson (1994), Spiegelhalter etc. (2004), Dmitrienko and Wang (2006). Appendix 9.5 contains a detailed derivation.

$$PP = \Phi\left\{\sqrt{\frac{2}{\frac{M_P}{N_P}+\frac{M_A}{N_A}}}\left[\sqrt{82}\frac{\bar{y}_{N_P}-\bar{x}_{N_A}}{\sqrt{2}\hat{\sigma}}-z_{0.95}\right]\right\} \quad (1)$$

where

- Φ stands for the cumulative distribution function for the standard Gaussian distribution. $\Phi(Z_{0.95}) = 0.95$.
- N_P and N_A denote the number of subjects observed at the interim analysis for placebo group and mirabegron group, respectively.
- M_P and M_A denote the number of subjects as yet unobserved at the interim analysis for placebo group and mirabegron, respectively. $M_P = 82 - N_P,\ M_A = 82 - N_A$.
- $\bar{y}_{N_P}$ and $\bar{x}_{N_A}$ are the LS means for placebo group and mirabegron group, respectively, and will be obtained from the ANCOVA model described in Section 6.4.1.1.
- $\hat{\sigma}$ is the estimated pooled standard deviation and will be obtained from the same ANCOVA model as described in Section 6.4.1.1.

Detailed information on the decision rule for futility and simulations illustrating the operational characteristics of such futility criteria can be found in Appendix 9.4.

The interim analysis will be conducted by an Independent Data Analysis Center (IDAC). An independent statistician at this IDAC will be unblinded to conduct the interim analysis.

Documentation of the interim analysis will be kept at the IDAC until after end of study treatment unblinding, and will then be transferred to the Study TMF.

The recommendation from the IDAC (to stop the study for futility or continue as planned) will be forwarded to the Therapeutic Area Head Medical Specialties, to act upon this

recommendation (see Appendix 9.6: Interim Analysis Report Template). His decision to stop or continue the study will be communicated to the study manager.

6.9.1 Data Sources

There are 2 databases: the IRT database provided by Cenduit and the clinical database. The IRT database will include all randomization and study drug supply data collected through IRT. The clinical database for the study is based on electronic data capture (EDC) technology. All clinical data should be captured electronically in the clinical database after the corresponding visits.

The unblinded randomization data and study drug supply data will only be located in the IRT database prior to the final database lock. The unblinded randomization data and study drug supply data will be transferred to the IDAC directly from the IRT database or IRT vendor for generating the Report. This process will maintain the confidentiality of the treatment codes by having only the IDAC statistician access the IRT independently of any Astellas personnel (e.g., clinical supply). This process will be documented by IRT vendor SOPs.

6.9.2 Data Cut-off and Data Extraction

The data cut-off date will be the target date for priority data cleaning. Data collected prior to data cut-off date will undergo standard data review and cleaning procedures by Signant Health and APGD Data Management. When priority data cleaning is deemed sufficient, the database will be extracted in SAS data format. All relevant (efficacy, disposition, study medication and demographic) data collected in the clinical database prior to the data extraction date will be included in the report.

6.9.3 Report Generation

The IDAC will develop analysis programs based on surrogate treatment codes prior to the cut-off date. The unblinded IDAC statistician that is not affiliated with APGD will receive the unblinded randomization data from the IRT. The IDAC statistician will replace surrogate treatment codes with actual treatment codes and generate outputs using the programs developed by the IDAC.

The IDAC statistician will prepare a statistical report, being the Interim Analysis Report (Appendix 9.6) and the TFL (Table, figure and listing) outputs, see below, for the Therapeutical Area Head Medical Specialties.

For the interim report the following tables will be presented:

Table Number	Table Title
1.1	Disposition Prior to Randomization, All Subjects With Informed Consent
1.2	Subject Classification, All Randomized Subjects
1.3	Subject Enrollment, Overall and by Country/Site, All Randomized Subjects
2	Demographics, Full Analysis Set
3	Results of ANCOVA of Change from Baseline in Mean Number of Micturitions per 24 hours at Week 12/EoT, Full Analysis Set

Note: no safety data will be presented

6.10 Additional Conventions

6.10.1 Analysis Windows

The study protocol gives the overall study schedule and the permissible intervals for these visits expressed as the number of days relative to visit 3/baseline. The total time each subject will be in the study will not exceed 18 weeks with a maximum of a screening period of 4 weeks and a maximum of 14 weeks for the investigational period.

Analyses will not exclude subject data due to the subject's failure to comply with the visit schedule.

All the assessments will be allocated to CRF visit based on Table 10.

Table 10 Visit Windows for Efficacy and Safety Parameters

CRF visit	Target day relative to Baseline (Visit 3, Day -1)	Analysis Window	Analysis visit
Bladder e-diary			
Visit 3	Last value prior to first dose	Day -7 to Day 1 (prior to first dose)	Baseline
Visit 4	Day 14	Day 7 to Day 21	Week 2
Visit 5	Day 28	Day 22 to Day 42	Week 4
Visit 6	Day 56	Day 43 to Day 70	Week 8
Visit 7	Day 84	Day 71 to Day 91	Week 12
EoT	Day 84	Day 1 to ≤ Last dose of double-blind study drug	EoT
Acceptability questionnaires			
Visit 7	Day 84	Day 71 to Day 91	Week 12
EoT	Day 84	Day 1 to ≤ Last dose of double-blind study drug	EoT
SBPM			
Visit 3	Last value prior to first dose	Day -7 to Day 1 (prior to first dose)	Baseline
Visit 4	Day 14	Day 7 to Day 21	Week 2
Visit 5	Day 28	Day 22 to Day 42	Week 4
Visit 6	Day 56	Day 43 to Day 70	Week 8
Visit 7	Day 84	Day 71 to Day 91	Week 12
EoT	Day 84	Day 1 to ≤ Last dose of double-blind study drug	EoT

Table continued on next page

CRF visit	Target day relative to Baseline (Visit 3, Day -1)	Analysis Window	Analysis visit
Vital signs			
Visit 1			Screening
Visit 3	Last value prior to first dose	Day -7 to Day 1 (prior to first dose)	Baseline
Visit 5	Day 28	Day 22 to Day 56	Week 4
Visit 7	Day 84	Day 57 to Day 91	Week 12
EoT	Day 84	Day 1 to ≤ Last dose of double-blind study drug	EoT
Visit 8[#]	Day 98	Day 92 or EoT + 1 to no more than 17 days after EoT	Week 14 (EoS)
ECG			
Visit 1			Screening
Visit 3	Last value prior to first dose	Day -7 to Day 1 (prior to first dose)	Baseline
Visit 5	Day 28	Day 22 to Day 56	Week 4
Visit 7	Day 84	Day 57 to Day 91	Week 12
EoT	Day 84	Day 1 to ≤ Last dose of double-blind study drug	EoT
Clinical Laboratory Tests (Hematology, Biochemistry and Urinalysis)			
Visit 1			Screening
Visit 3	Last value prior to first dose	Day -7 to Day 1 (prior to first dose)	Baseline
Visit 5	Day 28	Day 22 to Day 56	Week 4
Visit 7	Day 84	Day 57 to Day 91	Week 12
EoT	Day 84	Day 1 to ≤ Last dose of double-blind study drug	EoT
Visit 8[#]	Day 98	Day 92 or EoT to no more than 17 days after EoT	Week 14 (EoS)
Post Void Residual (PVR) Volume			
Visit 3	Last value prior to first dose	Day -7 to Day 1 (prior to first dose)	Baseline
Visit 5	Day 28	Day 22 to Day 56	Week 4
Visit 7	Day 84	Day 57 to Day 91	Week 12
EoT	Day 84	Day 1 to ≤ Last dose of double-blind study drug	EoT
Visit 8[#]	Day 98	Day 92 or EoT to no more than 17 days after EoT	Week 14 (EoS)

[#]Visit 8 may occur earlier due to early discontinuation as it is the last scheduled study visit.

For non-diary data, if a subject has more than one non-missing value within a visit window, the non-missing assessment which is closest to the target day within a window will be used. If two or more values are equally close and on different days, the latest non-missing value will be used. If two or more values are equally close and on the same day, the mean will be used for continuous variables or the worst observed case for categorical variables.

For diary data, the assessment date for the whole diary will be considered to be the date of the last valid day of the diary. If more than one diary has an assessment date within the same window, and if this results in more than one non-missing value of a diary variable, the non-missing value with the diary assessment day that is closest to the target day will be used. In case of ties on different days, the later non-missing value will be used. In case of ties located on the same side of the target day (i.e., more than one value for the same day), the mean of the values will be used for continuous variables and the worst value for categorical variables. For analyzing diary data the labels of the study visits will not be used, they will be assigned based on the dates of assessment.

A two-day window around the visit date will be applied for the assignment of study period for exposure data.

6.10.2 Imputation Rules for Incomplete Dates

Missing primary and secondary efficacy endpoints will be handled by Last Observation Carried Forward (LOCF) for continuous variables. More details are provided in the relevant sections for those endpoints (Sections 6.4.1 and 6.4.2).

As a general principle, no imputation of missing data for other variables will be done. Exceptions are the start dates of AEs and start and stop dates of concomitant medication. The imputed dates will be used to allocate the concomitant medication and AEs to a treatment group, in addition to determining whether an AE is/is not treatment emergent. Listings of the AEs and concomitant medications will present the actual partial dates; imputed dates will not be shown.
Values (visit 7/week 12) for certain safety variables (lab values, vital signs values and ECG) and the go-to-bed time for Sunday (eDiary) will be imputed, see below.
For calculation of percentiles the not collected day of birthday, in the eCRF, will be imputed.

6.10.2.1 Imputation Rules for Diary Dates

The go-to-bed time for Sunday is not captured in the diary. This time will be imputed as occurring a short time after the second (evening) SBPM was performed: time of second SBPM plus 1 hour. If there is no time entered in the diary for the evening SBPM, then the go-to-sleep time will be imputed as 23:59 (on Sunday), see Section 5.

6.10.2.2 Imputation Rules for Adverse Events

For AEs, a missing or incomplete onset date will be imputed according to the following conventions.

If an onset date is missing or only the year is known, the imputed onset date will be the date of first dose of study drug.

If only the year is known for the AE onset date, the imputed onset date will be the latest of the following non-missing dates:

- Date of first dose of study drug
- January 1 of the year of AE onset date

If only the month and year is known for the onset date, set the surrogate onset date to the first day of that month and then apply the following rules.

- If the month and year of the onset date is prior to the month and year of the first dose of study drug, then the surrogate onset date will be the imputed onset date.
- If the month and year of the onset date is on or after the month and year of the first dose of study drug, then the imputed onset date will be the **latest** of the following non-missing dates:
  - Date of first dose of study drug
  - Surrogate onset date

If the imputed onset date is after the adverse event end date, the imputed onset date will be the same as the adverse event end date.

6.10.2.3 Imputation Rules for Concomitant Medications

In case of missing partial start and stop dates for concomitant medications, the following rules will be used:

If the start date is missing or partial:

- if the month is missing, use January
- if the day is missing, use the first day of the month under consideration
- if the year is missing, use year of the informed consent date
- if the entire date is missing, use informed consent date

If the stop date is missing or partial:

- if the month is missing, use December
- if the day is missing, use the last day of the month under consideration
- if the year or the entire date is missing, set the stop date to December 31st, 2099

If the imputed start date is after the stop date, then the imputed start date will be 1 day prior to the stop date.

6.10.2.4 Imputation Rules for Missing Safety Variables (Lab values, Vital Signs, ECGs or PVR)

In case subject safety data of lab values, vital signs, PVR or ECGs at EoT (visit 7/week 12) is missing, the LOCF method will be used for the summaries or shifts from baseline to EoT only. No values at baseline and no data from more than 5 days after the last dose of study medication will be carried forward to post-baseline visits.

6.10.2.5 Imputation of Age in Months

For the calculation of percentiles of height and weight the age in months is required. Since we capture in the eCRF for birthday only the month and year and not the day, the 15th of the month will be imputed for day.

6.10.3 Outliers

All values will be included in the planned analyses. A sensitivity analysis excluding outliers may be performed as an additional secondary analysis, if considered appropriate by the study statistician or the medical expert.

7 REVISION AND RATIONALE

7.1 List of Changes in SAP Version 2.0 from Version 1.0

The changes from the approved SAP Version 1.0 (Dated 07-Aug-2020) to Version 2.0 that impact analyses are listed with the rationale in the table below.

SAP Section	Description	Rationale
2.1	Secondary Objective regarding PK analysis updated	To be consistent with protocol
	Exploratory Objective: age group is added and endpoint "mean number of micturitions per 24 hours" is added to list of endpoints.	To be consistent with protocol
2.3	Text added at the end of the paragraph	Text added for clarification
3	Text is added to state that the assumptions about sample size, in addition to the levels for Type-1 error and power, were agreed upon with Pediatric Committee in the current Pediatric Investigation Plan.	This update is made for clarification of assumptions.
5.4.1	Text regarding TEAEs updated to reflect text in protocol.	Text differed from the text in protocol.
	"In any of the treatment groups" is added to "Common TEAEs are defined as preferred terms (PTs) that have been reported by at least 5% of the subjects"	To be in agreement with other studies
5.5.2	Weight categories in the 2 tables are updated	To be consistent with protocol. The minimum weight of subjects is changed from 11 kg to 13 kg
6.4.1.4	Added section on scoring by caregiver and subject	To perform a descriptive analysis to investigate the effect of scores by caregiver and subject differ.
6.4.1.5	Added "Symptomatic UTI" as subgroup	To be consistent with protocol. This was added in version 2 of the protocol
6.4.3	Added Mean Number of Micturitions per 24 hours as exploratory endpoint	To be consistent with protocol. This endpoint was added in version 2 of the protocol
6.5.2.1	Added section on Central and Local Laboratories	Lab data can also come from local laboratories, version 2 of protocol.
6.5.7	Subgroup analyses also done for vital signs.	Vital signs was missing
	Added Symptomatic UTI as subgroup	To be consistent between efficacy and safety analyses
6.10.1	Table 10: Day 2 → Day 1	Subject may stop treatment on the day of the first administration of study drug
6.10.2	The time to-go-to sleep on Sunday is not captured on the eDiary	To describe here that this missing time will be imputed. Imputation also described in Section 5 of SAP

7.2 List of Changes in SAP Version 3.0 from Version 2.0

The changes from the approved SAP Version 2.0 (dated 24-Sep-2021) to Version 3.0 that impact analyses are listed with the rationale in the table below.

SAP Section	Description	Rationale
2.1	Secondary efficacy variable "Mean number of nighttime incontinence episodes per 24 hours" was deleted	Endpoint will no longer be analyzed due to changes in inclusion and exclusion criteria
	Exploratory Endpoints. The following exploratory endpoints are added: ● Mean number of micturitions per 24 hours, adjusted for fluid intake: o Number of micturitions per 24 hours that would be required at week 12/EoT to void the same total volume as voided at baseline o Difference in mean number of micturitions per 24 hours adjusted for fluid intake compared to the expected mean number ● Mean volume voided per micturition as a percentage of expected bladder capacity ● Categorized mean volume voided per micturition as a percentage of expected bladder capacity (results and changes from baseline).	To be in line with previous studies
	Added exploratory endpoint: "Change from baseline at the end of the 12-week treatment period (adolescents only) in Maximum Volume Voided"	To correct an omission. In previous version this end-point was only mentioned for children.
4.5.1	Eligibility Deviations: Inclusion criterion 16 has been removed	Not part of ICCS criteria; unduly restrictive for enrollment
	Exclusion criterion 1 has been removed	Not part of ICCS criteria; unduly restrictive for enrollment
	Exclusion criteria 3 and 27 have been removed	No reason to exclude these subjects as monosymptomatic enuresis is known to co-occur with OAB
5.4.2.3	For calculations reference is made to Center for Disease Control and Prevention website	Using validated SAS® code
5.2.1	Derivation of mean volume voided was corrected.	Derivation was not correct
Table continued next page		

SAP Section	Description	Rationale
5.2.5	Text in paragraph was updated	For clarification
5.4.1	An AE designated as serious by either the investigator or determined by Important Medical Event process will be summarized as an SAE. This paragraph was added with "However, if the event is on IME list but medical monitor agrees with investigator that event is not SAE, then event can be downgraded."	Sentence was added for further clarification regarding the process.
5.4.5 and 9.5	Deletion of text with respect to External ECG data and Abnormal ECGs	No external ECG data will be collected
6.2.3	Summaries for percentiles of height and weight, with respect to age and sex, will be provided using SAS code and dataset from Center for Disease Control and Prevention website	Using validated SAS® code.
6.4.1.3.4	Description for a (nonparametric) estimate for the difference between placebo and mirabegron, with associated (asymptotic) 90% CI was added	To present a nonparametric estimate of the treatment difference
6.4.1.4	Definition of valid diary for caregiver and for subject was added	For clarification and was missing in previous version
6.4.3	Text on mean number of micturitions per 24h adjusted for fluid intake was updated	For clarification
6.5.1	Text regarding "Sponsor's list of Always Serious terms" was updated by adding "or upgraded according to the Important Medical Event process".	New process: IME
6.5.3	Blood pressure will also be presented by age groups 5 - <8 and 8 - <12 years of age	As suggested by FDA
6.5.3.2	Calculation of SBPM values added	For clarification
6.5.6.2	Section of Body Weight and Height was added	Missing in previous version
6.10.1	Analysis window for PVR volume added in Table 10	Missing in previous version
	Table 10: for SBPM the row with EoS was removed and row (visit 8) was added for vital signs	No SBPM assessment planned for End of Study, but one for vital signs
6.10.2.4	Imputation of missing PVR Volume was updated	To be in line with other variables

Table continued next page

SAP Section	Description	Rationale
6.10.2.5	For birthday, only the month and year are captured, and not the day. If day is needed for calculations the 15th of the month will be imputed for day	Added because for some calculations the day of birthday is needed
8	List of references was updated	Two new references added and one removed.
9.1	Table which presents not allowed medication is deleted. Reference is made to Table 12.6 in protocol.	New table is presented in protocol
9.3	Appendix 9.3 Calculation of Z-score for height by age and sex has been removed	For calculation of Z-score the SAS code and dataset stored on website from Center for Disease Control and Prevention will be used
9.8 and 9.9	Names and titles updated	New titles and contributors/approvers

8 REFERENCES

2017 American Academy of Pediatric Clinical Practice Guidelines, Flynn JT, Kaelber DC, Baker-Smith CM, et al. Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents. Pediatrics.2017;140(3):e20171904)

Dmitrienko A. and Wang M.D. *Bayesian predictive approach to interim monitoring in clinical trials.* Statistical in Medicine, 2006;25:2178-2195.

Fleming S, Thompson M, Stevens R, Heneghan C, Plüddemann A, Maconochie I, et al. *Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies.* Lancet. 2011;377:1011-18.

Flynn JT, Kaelber DC, Baker-Smith CM, et al. *Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents.* Pediatrics. 2017;140(3):e20171904)

Geisser S. and Johnson W. Interim analysis for normally distributed observables. *Multivariate Analysis and Its Applications, IMS Lecture Notes – Monograph Series*, 1994;Volume 24, 263-279.

Hollander M, Wolfe DA. Nonparametric Statistical Methods. 2nd ed. John Wiley, 75-81.

ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

Karvanen J. The Statistical Basis of Laboratory Data Normalization. Drug Information Journal. Vol 37, 101-107, 2003.

Kliegman RM, Stanton BMD, St. Geme J, Schor NF, Behrman RE. Nelson Textbook of Pediatrics, 19th edition. Elsevier Science Health 2007

Marschall-Kehrel D, Feustel C, Persson de Geeter C, Stehr M, Radmayr C, Sillén U, et al. Treatment with propiverine in children suffering from nonneurogenic overactive bladder and urinary incontinence: results of a randomized placebo-controlled phase 3 clinical trial. Eur Urol. 2009;55:729-36.

Sender Herschorn, Christopher R. Chapple, Robert Snijder, Emad Siddiqui, Linda Cardozo. Could Reduced Fluid Intake Cause the Placebo Effect Seen in Overactive Bladder Clinical Trials? Analysis of a Large Solifenacin Integrated Database. Urology, April 2017:106

Spiegelhalter D.J., Abrams K.R., and Myles, J.P.. Bayesian Approaches to Clinical Trials and Health-care Evaluation. 2004, Chichester, England: John Wiley & Sons.

The Fourth Report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. 2005. U.S. Department of Health and Human Services, National Institutes of Health National Heart, Lung, and Blood Institute. NIH Publication No. 05-5267. Originally printed September 1996 (96-3790) Revised May 2005.
Available from: https://www.nhlbi.nih.gov/files/docs/resources/heart/hbp_ped.pdf

9 APPENDICES

9.1 Example of SAS Code for Nonparametric ANCOVA

As sensitivity analysis for the primary efficacy endpoint a nonparametric ANCOVA on rank-transformed data (changes from baseline) will be performed (stratified rank ANCOVA). The p-value of the treatment difference will be assessed using a rank ANCOVA stratified by region, see Section on Sensitivity Analysis for Primary Efficacy Endpoint.

The SAS code used to implement this test will be similar to that shown below:

```
** Ranks within each region will be derived across the two treatment groups for the baseline
value (base) and the change from baseline value (chg) stored in the SAS dataset called ranks;
PROC RANK   NPLUS1   TIES=MEAN   OUT=ranks   DATA=dataset;
     BY        Region;
     RANKS  Base_r   Chg_r;
     VAR       Base   Chg;
RUN;


** Linear regression models will be applied to these ranks by region;
PROC REG   DATA=ranks   NOPRINT;
     BY  Region;
     MODEL   Chg_r = Base_r   Sex ;
     OUTPUT  OUT = residual   R = Resid;
RUN;

** The p-value will be assessed using the stratified mean score test:
PROC FREQ    DATA=residual ;
     TABLES   Region*Sex*Trt*Resid /  CMH2   noprint;
     ODS   OUTPUT   CHM = outp;
RUN;
```

9.2 AEs of Interest not covered by Standard MedDRA SMQs latest Version

The following is a list of AE terms to programmatically flag subjects with AE's of interest (see Section 5.4.1).

Type	Term	Code	AE of Interest	MedDRA 24.0 or Higher Search Criteria
LLT	Acute retention of urine	10001055	Acute urinary retention	Selected LLT (non-SMQ)
PT	Residual urine volume	10050832	Urinary retention	Selected PT's (non-SMQ)
PT	Residual urine volume increased	10067758	Urinary retention	Selected PT's (non-SMQ)
PT	Urinary retention	10046555	Urinary retention	Selected PT's (non-SMQ)

Note: Most recent MedDRA version in which AEs are coded will be used.

9.3 2017 American Academy of Pediatric Clinical Practice Guidelines

From 2017 (American Academy of Pediatric) Clinical Practice Guidelines, Table 4, Males, see Reference Flynn, *et al*, 2017

Age (years)	BP Percentile	Systolic Blood Pressure (mmHg) Percentiles							Diastolic Blood Pressure (mmHg) Percentiles						
		5th	10th	25th	50th	75th	90th	95th	5th	10th	25th	50th	75th	90th	95th
5	Height (cm)	104.4	106.2	109.1	112.4	115.7	118.6	120.3	104.4	106.2	109.1	112.4	115.7	118.6	120.3
	90th	103	104	105	106	107	108	108	63	64	65	65	66	67	67
	95th	107	108	109	109	110	111	112	66	67	68	69	70	70	71
	95th+12mmHg	119	120	121	121	122	123	124	78	79	80	81	82	82	83
6	Height (cm)	110.3	112.2	115.3	118.9	122.4	125.6	127.5	110.3	112.2	115.3	118.9	122.4	125.6	127.5
	90th	105	105	106	107	109	110	110	66	66	67	68	68	69	69
	95th	108	109	110	111	112	113	114	69	70	70	71	72	72	73
	95th+12mmHg	120	121	122	123	124	125	126	81	82	82	83	84	84	85
7	Height (cm)	116.1	118	121.4	125.1	128.9	132.4	134.5	116.1	118	121.4	125.1	128.9	132.4	134.5
	90th	106	107	108	109	110	111	111	68	68	69	70	70	71	71
	95th	110	110	111	112	114	115	116	71	71	72	73	73	74	74
	95th+12mmHg	122	122	123	124	126	127	128	83	83	84	85	85	86	86
8	Height (cm)	121.4	123.5	127	131	135.1	138.8	141	121.4	123.5	127	131	135.1	138.8	141
	90th	107	108	109	110	111	112	112	69	70	70	71	72	72	73
	95th	111	112	112	114	115	116	117	72	73	73	74	75	75	75
	95th+12mmHg	123	124	124	126	127	128	129	84	85	85	86	87	87	87
9	Height (cm)	126	128.3	132.1	136.3	140.7	144.7	147.1	126	128.3	132.1	136.3	140.7	144.7	147.1
	90th	107	108	109	110	112	113	114	70	71	72	73	74	74	74
	95th	112	112	113	115	116	118	119	74	74	75	76	76	77	77
	95th+12mmHg	124	124	125	127	128	130	131	86	86	87	88	88	89	89
10	Height (cm)	130.2	132.7	136.7	141.3	146.9	150.1	152.7	130.2	132.7	136.7	141.3	146.9	150.1	152.7
	90th	108	109	111	112	113	115	116	72	73	74	74	75	75	76
	95th	112	113	114	116	118	120	121	76	76	77	77	78	78	78
	95th+12mmHg	124	125	126	128	130	132	133	88	88	89	89	90	90	90
11	Height (cm)	134.7	137.3	141.5	145.4	151.3	155.8	158.5	134.7	137.3	141.5	145.4	151.3	155.8	158.5
	90th	110	111	112	114	116	117	118	74	74	75	75	75	76	76
	95th	114	114	116	118	120	123	124	77	78	78	78	78	78	78
	95th+12mmHg	126	126	128	130	132	135	136	89	90	90	90	90	90	90
12	Height (cm)	140.3	143	147.5	152.7	157.9	162.6	165.5	140.3	143	147.5	152.7	157.9	152.6	165.5
	90th	113	114	115	117	119	121	122	75	75	75	75	75	76	76
	95th	116	117	118	121	124	126	128	78	78	78	78	78	79	79
	95th+12mmHg	128	129	130	133	136	138	140	90	90	90	90	90	91	91

From 2017 (American Academy of Pediatric) Clinical Practice Guidelines, Table 5, Females, see References

Age (years)	BP Percentile	Systolic Blood Pressure (mmHg) Percentiles							Diastolic Blood Pressure (mmHg) Percentiles						
		5th	10th	25th	50th	75th	90th	95th	5th	10th	25th	50th	75th	90th	95th
5	Height (cm)	103.6	105.3	108.2	111.5	114.9	118.1	120	103.6	105.3	108.2	111.5	114.9	118.1	120
	90th	104	105	106	107	108	109	110	64	65	66	67	68	69	70
	95th	108	109	109	110	111	112	113	68	69	70	71	72	73	73
	95th+12mmHg	120	121	121	122	123	124	125	80	81	82	83	84	85	85
6	Height (cm)	110	111.8	114.9	118.4	122.1	125.6	127.7	110	111.8	114.9	118.4	122.1	125.6	127.7
	90th	105	106	107	108	109	110	111	67	67	68	69	70	71	71
	95th	109	109	110	111	112	113	114	70	71	72	72	73	74	74
	95th+12mmHg	121	121	122	123	124	125	126	82	83	84	84	85	86	86
7	Height (cm)	115.9	117.8	121.1	124.9	128.8	132.5	134.7	115.9	117.8	121.1	124.9	128.8	132.5	134.7
	90th	106	106	107	109	110	111	112	68	68	69	70	71	72	72
	95th	109	110	111	112	113	114	115	72	72	73	73	74	74	75
	95th+12mmHg	121	122	123	124	125	126	127	84	84	85	85	86	86	87
8	Height (cm)	121	123	126.5	130.6	134.7	138.5	140.9	121	123	126.5	130.6	134.7	138.5	140.9
	90th	107	107	108	110	111	112	113	69	70	71	72	72	73	73
	95th	110	111	112	113	115	116	117	72	73	74	74	75	75	75
	95th+12mmHg	122	123	124	125	127	128	129	84	85	86	86	87	87	87
9	Height (cm)	125.3	127.6	131.3	135.6	140.1	144.1	146.6	125.3	127.6	131.3	135.6	140.1	144.1	146.6
	90th	108	108	109	111	112	113	114	71	71	72	73	73	73	73
	95th	112	112	113	114	116	117	118	74	74	75	75	75	75	75
	95th+12mmHg	124	124	125	126	128	129	130	86	86	87	87	87	87	87
10	Height (cm)	129.7	132.2	136.3	141	145.8	150.2	152.8	129.7	132.2	136.3	141	145.8	150.2	152.8
	90th	109	110	111	112	113	115	116	72	73	73	73	73	73	73
	95th	113	114	114	116	117	119	120	75	75	76	76	76	76	76
	95th+12mmHg	125	126	126	128	129	131	132	87	87	88	88	88	88	88
11	Height (cm)	135.6	138.3	142.8	147.8	152.8	157.3	160	135.6	138.3	142.8	147.8	152.8	157.3	160
	90th	111	112	113	114	116	118	120	74	74	74	74	74	75	75
	95th	115	116	117	118	120	123	124	76	77	77	77	77	77	77
	95th+12mmHg	127	128	129	130	132	135	136	88	89	89	89	89	89	89
12	Height (cm)	142.8	145.5	149.9	154.8	159.6	163.8	166.4	142.8	145.5	149.9	154.8	159.6	163.8	166.4
	90th	114	115	116	118	120	122	122	75	75	75	75	76	76	76
	95th	118	119	120	122	124	125	126	78	78	78	78	79	79	79
	95th+12mmHg	130	131	132	134	136	137	138	90	90	90	90	91	91	91

9.4 Futility Analysis Simulation Report

1.0 Background

This is a Phase 3 study of mirabegron in subjects with an Overactive Bladder (OAB). The study includes mirabegron and placebo. The primary efficacy endpoint is the change from baseline in mean number of micturitions per 24 hours at the end of the 12-week treatment period for children (week 12/EoT). A negative change from baseline is considered a beneficial effect. A total sample size of 184 evaluable subjects will be randomized in a 1:1 ratio to the two treatment groups. It is expected that 164 subjects will be evaluable for the primary analysis. For the calculations it is assumed that 82 subjects per arm are evaluable.

2.0 Interim Evaluation

One comparative interim evaluation is planned to take place when 50%, i.e., 41 per treatment arm, of planned subjects to be randomized have had their week 12/EoT assessment. The intent of planning and conducting an interim efficacy analysis is to evaluate whether the mirabegron group meets futility criteria.

At the interim evaluation, a Bayesian predictive probability for the event of a final 1-sided p-value ≤ 0.05 will be calculated, comparing the mirabegron group with placebo. If the predictive probability for the interim evaluation is ≤ 0.05 (i.e., there is a $\leq$5% chance of obtaining a statistically significant result for the primary efficacy analysis for mirabegron group at the final analysis), the mirabegron will be considered to have met futility criteria.

2.1 Decision Rule for Futility

At the interim evaluation, the following predictive probability will be calculated:

Predictive Probability of ({final 1-sided p-value of mirabegron vs. placebo ≤ 0.05} | Data).

Given the data at the interim analysis, a predictive probability of a positive study will be calculated. A positive study is defined as the final 1-sided p-value ≤ 0.05 when comparing mirabegron vs. placebo in the primary efficacy analysis. If the predictive probability is no more than 0.05 at this interim evaluation, the study will be recommended to stop for futility. Otherwise, the study will continue.

2.2 Calculation of Predictive Probability

At the interim evaluation, the predictive probability is expressed as an integration of the probability of {final one-sided p-value of mirabegron vs. placebo ≤ 0.05} over the posterior distribution of the treatment difference given the observed data:

$$\int \Pr(\{\text{Final one-sided } p\text{ - value} \leq 0.05\}|\theta) \times \Pr(\theta|\text{data})d\theta,$$

where θ stands for the difference in mean change from baseline to week 12/EoT in mean number of micturitions per 24 hours between mirabegron group and placebo group. Such predictive probability can be computed through simulation.

Since we use non-informative prior distribution of difference in mean change from baseline to week 12/EoT in mean number of micturitions per 24 hours between mirabegron group and

placebo group, it is assumed that the mean number of micturitions per 24 hours follows a Normal distribution and equal sample size per arm at the interim evaluation, the Bayesian predictive probability for the event of a final 1-sided p-value of mirabegron vs. placebo ≤ 0.05 (denoted by *PP*) can be calculated using a closed form based on the methods described in Geisser and Johnson (1994), Dmitrienko and Wang (2006).

$$PP = \Phi\left[\frac{1}{\sqrt{M}}\left(\sqrt{N+M}\,Z_N - \sqrt{N}\,Z_{0.95}\right)\right] \qquad (1)$$

where

Φ stands for the cumulative distribution function for the standard normal distribution;

N = number of subjects per arm observed at the interim evaluation, here N=41;

M = number of subjects per arm as yet unobserved at the interim evaluation, here M = 82 – N

Z_N = test statistic, where the numerator (difference in mean change between mirabegron and placebo) will be obtained from the sample means of the mirabegron group and placebo group,

$$Z_N = \frac{\text{Difference in LS mean change (mirabegron} - \text{placebo)}}{\sigma\sqrt{\frac{2}{N}}} \qquad (2)$$

The least squares (LS) means for placebo group and mirabegron group will be obtained from the ANCOVA with treatment (mirabegron and placebo), visit, geographical region, sex, and treatment by visit interaction as fixed effects and baseline value (mean number of micturitions per 24 hours at baseline) as covariate. The estimated pooled standard deviation (σ) will also be obtained from the same ANCOVA model.

3.0 Simulation Algorithm

Simulations were conducted to evaluate the operating characteristics of this design, including the probability of meeting the futility criteria at the interim evaluation, and the probability of continuing to the end and having final 1-sided p-value ≤ 0.05.

Change from baseline to week 12 in mean number of micturitions per 24 hours for each subject is generated from a Normal distribution. The common standard deviation for mirabegron group and placebo group is fixed at 2.3 (see Section 3, Sample Size). The planned sample size is 82 subjects per arm. The scheduled interim evaluation is the moment when 82 subjects (i.e., 41 subjects per arm) have their week 12/EoT mean number of micturitions per 24 hours. The outcome of the simulation is based on 50,000 runs (i.e., repeated trials).

Six (6) scenarios were evaluated. Each scenario specifies a mean change from baseline to week 12 in mean number of micturitions per 24 hours in the mirabegron group, while it keeps the mean change from baseline to week 12 in mean number of micturitions per 24 hours for placebo group fixed as –0.9.

See Table 7.1-1 for the six scenarios, where δ denotes the treatment difference between mirabegron group and placebo group in mean change from baseline to week 12 mean number of micturitions per 24 hours and a negative value of δ indicates a favorable treatment effect, and μ_A denotes the corresponding mean change from baseline to week 12 in mean number of micturitions per 24 hours for mirabegron group.

Table 7.1-1: Simulation Scenarios

Scenario #	1	2	3	4	5	6
CfB for mirabegron	-0.9	-1.1	-1.3	-1.5	-1.7	-1.8
δ*	0.0	-0.2	-0.4	-0.6	-0.8	-0.9

CfB: Change from Baseline

*: Treatment difference (δ) = mirabegron – placebo. Negative value of δ indicates favorable treatment effect. σ = 2.3 for both treatment groups.

The simulation steps are as follows:

1. Generate 41 observations from a Normal distribution with mean = 0.0 and standard deviation = 2.3 for placebo group and 41 observations from a Normal distribution with mean δ (any scenario in Table 7.1-1) and standard deviation = 2.3 for mirabegron group.
2. Obtain the treatment difference between mirabegron group and placebo group using mean difference (mirabegron minus placebo) as fixed effect described in formula (2) based on the generated data.
3. Apply formula (1) using σ = 2.3 to calculate the predictive probability of {final one-sided p-value of mirabegron vs. placebo ≤ 0.05}, where for the interim evaluation N = 41, and M = 41.
4. If the futility criteria are not met, then generate for each treatment group data for 41 subjects to complete the trial.
5. Repeat Step 1-4 50,000 times (i.e., 50,000 simulated trials) for each scenario. Count the number of simulated trials which meet the futility criterion. In simulation, as long as mirabegron group meets the futility criterion, the simulated trial will be stopped. The probability of meeting the futility criterion for a given scenario is the percentage of the trials meeting the futility criterion out of the 50,000 simulated trials. Likewise, the probability of success is the percentage of the trials continuing to the end and having final 1-sided p-value of mirabegron vs. placebo ≤ 0.05.

The above simulation procedure was performed for a number of treatment difference scenarios. The simulations were conducted using the software package SAS®.

3.1 Operating Characteristics

Table 7.1-2 presents the probabilities of futility at the interim evaluation, as well as the probability of success (defined as continuing to the end and having final 1-sided p-value of mirabegron vs. placebo ≤ 0.05), for each treatment difference scenario. In each treatment

difference scenario, we assume the common standard deviation of mean change from baseline to week 12/EoT in mean number micturitions per 24 hours is 2.3.

Table 7.1-2: Probability of Futility and Success, σ = 2.3

Treatment difference mirabegron vs. placebo (δ)*	*Pr* (Futility)	*Pr* (Success)
0.0	0.501	0.050
-0.2	0.345	0.134
-0.4	0.217	0.289
-0.6	0.118	0.503
-0.8	0.057	0.713
-0.9	0.038	0.797

* Negative value of δ indicates favorable treatment effect.

When there is no treatment difference between the mirabegron and placebo group ($\delta = 0$), study success would be a false positive one. Table 7.1-2 shows that in 50.1% of the simulations the trial would stop for futility. Furthermore, 5.0% of the simulated trials continued to the end and had a 1-sided *p*-value ≤ 0.05 for the test to compare the mean changes between the mirabegron and placebo group. Therefore, the false positive rate for this scenario is 5.0%.

When the treatment difference is -0.9 mean number of micturitions per 24 hours, favoring mirabegron over placebo, the study would only be stopped for futility with a probability of 3.8%, whereas 79.7% of the simulated trials continued to the end and had a 1-sided p-value ≤ 0.05.

In addition, a sensitivity analysis for the assumption regarding the common standard deviation of mean change from baseline to week 12 in mean number of micturitions per 24 hours was performed on the same six treatment difference scenarios as described in Table 7.1-1 to assess the robustness of the design. The probabilities of futility and success assuming $\sigma = 2.5$ for each of the treatment difference scenarios are presented in Table 7.1-3.

For the efficacious scenarios where mirabegron shows favorable treatment effect over placebo ($\delta < 0$), the probabilities of futility are relatively higher and the probabilities of success are relatively lower compared to the results assuming $\sigma = 2.3$. For example, when the treatment difference is -0.6 points favoring mirabegron over placebo, the probability of futility in this scenario is 14.3% as compared to 11.8% assuming $\sigma = 2.3$, and the probability of success is 44.6%, as compared to 50.3% assuming $\sigma = 2.3$.

Table 7.1-3: Probability of Futility and Success, σ = 2.5

Treatment difference mirabegron vs. placebo (δ)*	*Pr* (Futility)	*Pr* (Success)
0.0	0.501	0.052
-0.2	0.357	0.127
-0.4	0.234	0.260
-0.6	0.143	0.446
-0.8	0.073	0.646
-0.9	0.051	0.739

* Negative value of δ indicates favorable treatment effect.

9.5 Derivation of Formula to Calculate Bayesian Predictive Probability

Derivation of Formula (1) as presented in Section 6.9 is described below.

Suppose two arms, placebo group and mirabegron group. The following notations are used:

- μ_P and μ_A denote the mean change from Baseline to Week 12/EoT in mean number of micturitions per 24 hours for the placebo and mirabegron group, respectively.
- σ denotes the common standard deviation.
- In total 164 subjects will be randomized to each of the two arms and will be in the FAS. Let NN denotes the total sample size per arm planned for the study to have a baseline and a Week 12/EoT value for the mean number of micturitions per 24 hours, $NN = 82$.
- N_P and N_A denote the number of subjects observed, with data, at the interim analysis for the placebo and mirabegron group, respectively.
- M_P and M_A denote the number of subjects as yet unobserved at the interim analysis for the placebo and mirabegron group, respectively. $M_P = NN - N_P = 82 - N_P,\ M_A = NN - N_A = 82 - N_A$.

Suppose that:

- For placebo group: $Y_1, Y_2, \cdots, Y_{N_P}, Y_{N_P+1}, \cdots, Y_{NN} \sim i.i.d.N(\mu_P, \sigma^2)$.
- For mirabegron group: $X_1, X_2, \cdots, X_{N_A}, X_{N_A+1}, \cdots, X_{NN} \sim i.i.d.N(\mu_A, \sigma^2)$.

At the interim analysis when N_P and N_A subjects have their Week 12/EoT data for placebo group (i.e., $y_1, y_2, \cdots, y_{N_P}$ have been observed) and mirabegron group (i.e., $x_1, x_2, \cdots, x_{N_A}$ have been observed), respectively, the posterior distribution of μ_P and μ_A assuming non-informative priors are:

$$\mu_P | y_1, y_2, \cdots, y_{N_P} \sim N\left(\bar{y}_{N_P}, \frac{\sigma^2}{N_P}\right), \text{where } \bar{y}_{N_P} = \frac{1}{N_P}\sum_{i=1}^{N_P} y_i.$$

$$\mu_A | x_1, x_2, \cdots, x_{N_A} \sim N\left(\bar{x}_{N_A}, \frac{\sigma^2}{N_A}\right), \text{where } \bar{x}_{N_A} = \frac{1}{N_A}\sum_{i=1}^{N_A} x_i.$$

For placebo group, the mean of as yet unobserved future data $Y_{N_P+1}, Y_{N_P+2} \cdots, Y_{NN}$ is denoted as $\bar{Y}_{M_P}$, $\bar{Y}_{M_P} = \frac{1}{M_P}\sum_{i=N_P+1}^{NN} Y_i$. The expectation and variance of $\bar{Y}_{M_P}$ given observed data $y_1, y_2, \cdots, y_{N_P}$ can be expressed as follows

$$E\left(\bar{Y}_{M_P} | data\right) = E\left(E\left(\bar{Y}_{M_P} | \mu_P\right) | data\right) = E(\mu_P | data) = \bar{y}_{N_P}$$

$$Var\left(\bar{Y}_{M_P} | data\right) = Var\left(E\left(\bar{Y}_{M_P} | \mu_P\right) | data\right) + E\left(Var\left(\bar{Y}_{M_P} | \mu_P\right) | data\right)$$

$$= Var(\mu_P | data) + E\left(\frac{\sigma^2}{M_P}\right) = \frac{\sigma^2}{N_P} + \frac{\sigma^2}{M_P} = \sigma^2\left(\frac{1}{N_P} + \frac{1}{M_P}\right)$$

Similarly for mirabegron group, the mean of as yet unobserved future data $X_{N_A+1}, X_{N_A+2}, \cdots, X_{NN}$ is denoted as $\bar{X}_{M_A}$, $\bar{X}_{M_A} = \frac{1}{M_A}\sum_{i=N_A+1}^{NN} X_i$. The expectation and variance of $\bar{X}_{M_A}$ given observed data $x_1, x_2, \cdots, x_{N_A}$ can be expressed as follows

$$E\left(\bar{X}_{M_A}|data\right) = E\left(E\left(\bar{X}_{M_A}|\mu_A\right)|data\right) = E(\mu_A|data) = \bar{x}_{N_A}$$

$$Var\left(\bar{X}_{M_A}|data\right) = Var\left(E\left(\bar{X}_{M_A}|\mu_A\right)|data\right) + E\left(Var\left(\bar{X}_{M_A}|\mu_A\right)|data\right)$$

$$= Var(\mu_A|data) + E\left(\frac{\sigma^2}{M_A}\right) = \frac{\sigma^2}{N_A} + \frac{\sigma^2}{M_A} = \sigma^2\left(\frac{1}{N_A} + \frac{1}{M_A}\right)$$

Hence the predictive distribution of $\bar{Y}_{M_P}$ and $\bar{X}_{M_A}$ given observed data are as follows

$$\bar{Y}_{M_P}|data \sim N\left(\bar{y}_{N_P}, \sigma^2\left(\frac{1}{N_P} + \frac{1}{M_P}\right)\right) \quad (1)$$

$$\bar{X}_{M_A}|data \sim N\left(\bar{x}_{N_A}, \sigma^2\left(\frac{1}{N_A} + \frac{1}{M_A}\right)\right) \quad (2)$$

Using (1) and (2), we have

$$E\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}|data\right) = M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}$$

$$Var\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}|data\right) = M_P^2 Var\left(\bar{Y}_{M_P}|data\right) + M_A^2 Var\left(\bar{X}_{M_A}|data\right)$$

$$= \sigma^2 M_P^2\left(\frac{1}{N_P} + \frac{1}{M_P}\right) + \sigma^2 M_A^2\left(\frac{1}{N_A} + \frac{1}{M_A}\right)$$

$$= \sigma^2\left[M_P^2\left(\frac{N_P + M_P}{N_P M_P}\right) + M_A^2\left(\frac{N_A + M_A}{N_A M_A}\right)\right]$$

$$= \sigma^2\left(M_P^2\frac{NN}{N_P M_P} + M_A^2\frac{NN}{N_A M_A}\right)$$

$$= \sigma^2 NN\left(\frac{M_P}{N_P} + \frac{M_A}{N_A}\right)$$

Hence, the predictive distribution of $\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}\right)$ given observed data is

$$M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}|data \sim N\left(M_P\bar{y}_{N_P} - M_A\bar{x}_{N_A}, \sigma^2 NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)\right) \quad (3)$$

We are interested in predicting whether the as yet unobserved future data together with the observed data will result in a one-sided p-value for the null hypothesis(H_0): $\mu_A - \mu_P \geq 0$, and the alternative hypothesis (H_1): $\mu_A - \mu_P < 0$. The standard test for testing H_0 versus H_1 at significance level $\alpha = 0.05$ is to reject H_0 if

$$\frac{\bar{Y}_{NN} - \bar{X}_{NN}}{\sigma\sqrt{\frac{2}{NN}}} > z_{0.95} \quad (4)$$

Now we compute the predictive probability of the rejection set in (4), which is denoted as *PP*.

$$PP = \Pr\left\{\frac{\bar{Y}_{NN} - \bar{X}_{NN}}{\sigma\sqrt{\frac{2}{NN}}} > z_{0.95}\right\}$$

$$= \Pr\left\{\bar{Y}_{NN} - \bar{X}_{NN} > z_{0.95}\sigma\sqrt{\frac{2}{NN}}\right\}$$

$$= \Pr\left\{\frac{N_P\bar{y}_{N_P} + M_P\bar{Y}_{M_P}}{NN} - \frac{N_A\bar{x}_{N_A} + M_A\bar{X}_{M_A}}{NN} > z_{0.95}\sigma\sqrt{\frac{2}{NN}}\right\}$$

Note that the observed $\bar{y}_{N_P}$and $\bar{x}_{N_A}$are fixed, but the as yet unobserved $\left(\bar{Y}_{M_P} - \bar{X}_{M_A}\right)$ are random following the predictive distribution specified in (3). Therefore,

$$PP = \Pr\left\{\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}\right) + \left(N_P\bar{y}_{N_P} - N_A\bar{x}_{N_A}\right) > z_{0.95}\sigma NN\sqrt{\frac{2}{NN}}\right\}$$

$$= \Pr\left\{M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A} > z_{0.95}\sigma\sqrt{2NN} - \left(N_P\bar{y}_{N_P} - N_A\bar{x}_{N_A}\right)\right\}$$

$$= \Pr\left\{\frac{\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}\right) - \left(M_P\overline{y}_{N_P} - M_A\overline{x}_{N_A}\right)}{\sigma\sqrt{NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)}} > \frac{z_{0.95}\sigma\sqrt{2NN} - \left(N_P\bar{y}_{N_P} - N_A\bar{x}_{N_A}\right) - \left(M_P\overline{y}_{N_P} - M_A\overline{x}_{N_A}\right)}{\sigma\sqrt{NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)}}\right\}$$

$$= \Pr\left\{\frac{\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}\right) - \left(M_P\overline{y}_{N_P} - M_A\overline{x}_{N_A}\right)}{\sigma\sqrt{NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)}} > \frac{z_{0.95}\sigma\sqrt{2NN} - NN\left(\bar{y}_{N_P} - \bar{x}_{N_A}\right)}{\sigma\sqrt{NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)}}\right\}$$

$$= \Pr\left\{\frac{\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}\right) - \left(M_P\bar{y}_{N_P} - M_A\bar{x}_{N_A}\right)}{\sigma\sqrt{NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)}} > \frac{z_{0.95}\sigma\sqrt{2} - \sqrt{NN}\left(\bar{y}_{N_P} - \bar{x}_{N_A}\right)}{\sigma\sqrt{\frac{M_A}{N_A} + \frac{M_P}{N_P}}}\right\}$$

From (3) we have

$$M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}|data \sim N\left(M_P\bar{y}_{N_P} - M_A\bar{x}_{N_A}, \sigma^2 NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)\right)$$

$$\text{Let } Z = \frac{\left(M_P\bar{Y}_{M_P} - M_A\bar{X}_{M_A}\right) - \left(M_P\bar{y}_{N_P} - M_A\bar{x}_{N_A}\right)}{\sigma\sqrt{NN\left(\frac{M_A}{N_A} + \frac{M_P}{N_P}\right)}}, \quad Z \sim N(0,1)$$

Hence,

$$PP = \Pr\left\{Z > \frac{z_{0.95}\sigma\sqrt{2} - \sqrt{NN}\left(\bar{y}_{N_P} - \bar{x}_{N_A}\right)}{\sigma\sqrt{\frac{M_A}{N_A} + \frac{M_P}{N_P}}}\right\}$$

$$= \Pr\left\{Z > \sqrt{\frac{2}{\frac{M_A}{N_A} + \frac{M_P}{N_P}}}\left[z_{0.95} - \sqrt{NN}\frac{\bar{y}_{N_P} - \bar{x}_{N_A}}{\sqrt{2}\sigma}\right]\right\}$$

$$= 1 - \Pr\left\{Z \leq \sqrt{\frac{2}{\frac{M_A}{N_A} + \frac{M_P}{N_P}}}\left[z_{0.95} - \sqrt{NN}\frac{\bar{y}_{N_P} - \bar{x}_{N_A}}{\sqrt{2}\sigma}\right]\right\}$$

$$= 1 - \Phi\left\{\sqrt{\frac{2}{\frac{M_A}{N_A} + \frac{M_P}{N_P}}}\left[z_{0.95} - \sqrt{NN}\frac{\bar{y}_{N_P} - \bar{x}_{N_A}}{\sqrt{2}\sigma}\right]\right\}$$

$$= \Phi\left\{\sqrt{\frac{2}{\frac{M_A}{N_A} + \frac{M_P}{N_P}}}\left[\sqrt{NN}\frac{\bar{y}_{N_P} - \bar{x}_{N_A}}{\sqrt{2}\sigma} - z_{0.95}\right]\right\}$$

By substituting NN by 82, M_A by $82 - N_A$, M_P by $82 - N_P$, and σ by $\hat{\sigma}$, we have

$$PP = \Phi\left\{\sqrt{\frac{2}{\frac{82-N_A}{N_A}+\frac{82-N_P}{N_P}}}\left[\sqrt{82}\frac{\bar{y}_{N_P}-\bar{x}_{N_A}}{\sqrt{2}\hat{\sigma}}-z_{0.95}\right]\right\}$$

where

- Φ stands for the cumulative distribution function for the standard normal distribution.
- $\bar{y}_{N_P}$ and $\bar{x}_{N_A}$ are the LS means for placebo group and mirabegron group, respectively, and will be obtained from the ANCOVA model (using LOCF), described in Section 6.4.1.1 of the SAP.
- $\hat{\sigma}$ is the estimated pooled standard deviation and will be obtained from the same ANCOVA model as described above.

Note that when the numbers of subjects observed at the interim analysis for the mirabegron and placebo group are equal (i.e., $N_A = N_P = N$), let $M = NN - N = 82 - N$, the formula can be simplified to

$$PP = \Phi\left\{\sqrt{\frac{1}{M}}\left[\sqrt{82}\frac{\bar{y}_{N_P}-\bar{x}_{N_A}}{\sqrt{\frac{2}{N}}\hat{\sigma}}-\sqrt{N}z_{0.95}\right]\right\},$$

which is the same as the one used in the Appendix 9.4.

9.6 Interim Analysis Report Template

Study 178-CL-204 Interim Analysis

Futility Analysis Results

Generated by *[IDAC Statistician's Full Name]* on *ddMMMyyyy*

The performed Interim Analysis should determine if the chance of a positive study with respect to the primary endpoint at the final analysis is high enough to justify continuation of the study (Predictive Probability is ≥ 0.05); otherwise, the study should be stopped for futility.

Based on the following statistics calculated using the data by the cut-off date *ddMMMyyyy*:

- The number of subjects observed at the interim analysis for placebo and mirabegron group are *XX* and *XX*, respectively.
- The least squares means for placebo group and mirabegron group are *XX.X* and *XX.X*, respectively, for the mean change from baseline in mean number of micturitions per 24 hours at week12/EoT. The results are obtained from an ANCOVA model with treatment group, sex and geographical region as fixed effects and the mean number of micturitions per 24 hours at baseline as covariate. This analysis was performed with imputation of missing visit 7/week 12 data using the last observation carried forward (LOCF) method.
- The estimated pooled standard deviation is *XX.X*, obtained from the same ANCOVA model as described above.

Predictive Probability of ({final one-sided p-value of mirabegron vs. placebo ≤ 0.05}|Data) = 0.*XXX*

Hence, the study *[did or did not]* meet the futility criteria "Predictive Probability is ≥ 0.05". The IDAC recommends the study to *[continue or stop]*.

9.7 Key Contributors

Name, Degree, Title And Function	Department
PPD	Labcorp
PPD	Data Science
PPD	Clinical Operations
PPD	Therapeutical Area Medical Specialties
PPD	Therapeutical Area Medical Specialties
PPD	PAREXEL
PPD	Medical Specialties TA, Pharmacovigilance
PPD	Data Science
PPD	Labcorp

9.8 Author and Approver Signatures

Signatures

Prepared by: Date:

PPD

Date (DD MMM YYYY)

Approved by: *E-signatures are attached at end of document* Date:

PPD

Date (DD MMM YYYY)

Approved by: *E-signatures are attached at end of document* Date:

PPD

Date (DD MMM YYYY)