FERRING
PHARMACEUTICALS

Statistical Analysis Plan

Title of trial: A Randomised, Double-blind, Placebo-controlled, Response-adaptive Dose-finding Trial Investigating the Efficacy, Safety and Tolerability of Oral Doses of FE 201836, with Desmopressin Orally Disintegrating Tablet as a Benchmark, During 12 Weeks of Treatment for Nocturia due to Nocturnal Polyuria in Adults
NCT number: NCT03201419
Sponsor trial code: 000233
Date: 12 Dec 2019

STATISTICAL ANALYSIS PLAN

A Randomised, Double-blind, Placebo-controlled, Response-adaptive Dose-finding Trial Investigating the Efficacy, Safety and Tolerability of Oral Doses of FE 201836, with Desmopressin Orally Disintegrating Tablet as a Benchmark, During 12 Weeks of Treatment for Nocturia due to Nocturnal Polyuria in Adults

000233 (DAWN)

Investigational Product:	FE 201836 Oral Solution
Indication:	Nocturia due to Nocturnal Polyuria
Phase:	2
Author:	[redacted]
Date of issue:	
Version:	1.0

Change log

Version No.	Effective Date	Reason for the Change / Revision	Supersedes

Review and Approval of Statistical Analysis Plan

Author [redacted]

_____

[redacted]
[redacted] Statistician, Global Biometrics
Ferring Pharmaceuticals A/S

Reviewers [redacted]

_____

[redacted]
Statistical Programmer, Global Biometrics
Ferring Pharmaceuticals A/S

_____

[redacted]
[redacted] Pharmacovigilance physician, Global Pharmacovigilance, [redacted]
Ferring Pharmaceuticals A/S

_____

[redacted]
Medical writer, Medical Writing
Ferring Pharmaceuticals A/S

_____

[redacted]
[redacted] Statistician, Global Biometrics
Ferring Pharmaceuticals A/S

_____

[redacted]
Ferring Pharmaceuticals A/S

Approval by [redacted]

_____

[redacted]
[redacted], Biometrics
Ferring Pharmaceuticals A/S

TABLE OF CONTENTS

List of Tables

1 Introduction

1.1 General Considerations

This document describes the planned statistical analyses for 000233 DAWN. This SAP is based on the consolidated protocol Ver.5.0, dated 05 July 2018, with changes according to Global Amendments #01, #02, and #03 implemented.

1.2 Definitions/ Abbreviations

1.2.1 Definition of Terms

Terms	Definitions
Aggregated Average	Average over post-randomization visits up to and including the current visit
Baseline-ENR	Baseline measurement for the enrichment period, assessed at Visit 2
Baseline-RT	Baseline measurement for the randomized-treatment period, assessed at Visit 4
Δ_d	Effect of dose *d* of FE 201836 compared to placebo
e-Diary	Electronic diary filled out by the participant at home, covering Insomnia Severity Index (ISI), Nocturia Impact Diary (NI diary), Patient Global Impression of Severity (PGI-S), Patient Global Impression of Improvement (PGI-I), HSU 5 point Likert bother scale (Bother), Voiding Diary, and IMP intake
ED_{85}	The minimum dose most likely to have more than 85% of the effect of the maximum dose
Enrichment Period	Period between dispense of single-blind treatment and randomization
FUSP	First Undisturbed Sleep Period (FUSP) is defined as the time in minutes from the time of going to bed to the time of first nocturnal void, or time of awakening if no void occurred
MED	The minimally effective dose defined as the lowest dose that has at least a 0.3 larger reduction in mean nocturnal voids compared to placebo
μ_d	The mean dose-response at dose *d* of FE 201836 corresponding to two nocturnal voids at baseline-RT
Nocturnal diuresis rate	The nocturnal diuresis rate is the ratio of Nocturnal Urine Volume (NUV, mL) to total time in bed (min)
NPi	Nocturnal Polyuria index (NPi) is calculated as the ratio of the nocturnal urine volume (NUV) by the 24-hour urine volume
NUV	Nocturnal Urine Volume (NUV) is the total urine volume from 5 minutes after bedtime with the intention to sleep until the first void within 30 minutes of rising in the morning
Pr(dose *d* is ED85)	Posterior probability that the dose *d* is the ED_{85}
Pr(dose *d* is MED)	Posterior probability that the dose *d* is the MED
$Pr(\Delta_d <= -0.3)$	Posterior probability that the dose *d* has at least a 0.3 larger reduction in mean nocturnal voids compared to placebo
$Pr(\mu_{500} < \mu_{Plb})$	Posterior probability that 500 µg FE 201836 is superior to placebo
Safety-ENR	Safety analysis set for the enrichment period
Safety-RT	Safety analysis set for the randomized-treatment period

1.2.2 Abbreviations

Abbreviations	Meaning of abbreviations in document
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomic Therapeutic Chemical
CV	Coefficient of Variation
DMC	Data Monitoring Committee
ECG	Electrocardiogram
ENR	Enrichment Period
FUSP	First Undisturbed Sleep Period
HLGT	High Level Group Term
HLT	High Level Term
IMP	Investigational Medicinal Product
ISI	Insomnia Severity Index
ITT	Intention-To-Treat
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measurements (repeated-measurements ANCOVA)
NI	Nocturia Impact
NPi	Nocturnal Polyuria index
NUV	Nocturnal Urine Volume
ODT	Orally Disintegrating Tablet
PP	Per-Protocol
PT	Preferred Term
RT	Randomized-treatment Period
SAE	Serious Adverse Event
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event

1.3 Presentation of Results

1.3.1 Presentation of Descriptive Results

Categorical data will be summarized, unless otherwise stated, using frequency (n) and relative frequencies as percentage (i.e. n/N*100).

Continuous data will be presented for observed values and change form baseline for each visit and at last visit, unless otherwise stated, using the number of subjects (N), mean, standard deviation (SD), median, lower and upper quartiles, minimum, and maximum.

Only table shells for inferential statistics will be provided. End-of-Text and In-Text tables of descriptive statistics follow in-house standards.

1.3.2 Presentation of Inferential Results

Frequentist inferential results will be presented using the respective point estimates with two-sided 95% confidence intervals unless otherwise stated. Bayesian inferential results are expressed in terms of posterior means and medians and 95% credibility intervals, unless otherwise stated.

Generic table shells per type of analysis will be provided.

1.3.3 Subject Data Listings by Domain

All subject data will be presented by listings organized in accordance to ICH –E3. More precisely the following listings will be provided:

- Discontinued subjects
- Major protocol deviations
- Subjects (and observations) excluded from the efficacy analysis
- Demographic data
- Compliance
- Individual efficacy response data
- Adverse event listings (each subject)
- Listing of individual laboratory measurements by subject, when required by regulatory authorities

1.3.4 Subject Profiles

Subject profiles will be created using in-house standard SAS macros.

2 Trial Objectives and Endpoints

2.1 Objectives

Primary Objective

- To establish the dose-response of FE 201836 with respect to the number of nocturnal voids in subjects with nocturia due to NP.

Secondary Objectives

- To evaluate responder rates with regards to changes in number of nocturnal voids.
- To psychometrically validate the Nocturia Impact Diary© (NI Diary).
- To evaluate the subject benefit of FE 201836 based on the NI Diary data.
- To evaluate the clinical benefit of FE 201836 based on reduction in nocturnal voiding.
- To evaluate FE 201836 with respect to sleep benefit, i.e., duration of the First Undisturbed Sleep Period (FUSP) and sleep related Patient Reported Outcomes (PROs).
- To evaluate the PD effect of FE 201836 with respect to nocturnal diuresis rate and NP.

Safety Objective

- To evaluate the safety profile of FE 201836.

Exploratory Objectives

- To explore the biomarker copeptin, an AVP surrogate, to identify low nocturnal vasopressin levels in plasma.
- To explore the dose-response of FE 201836 with respect to total sleep time.
- To explore the association between total sleep time and FUSP.
- To explore the effect of FE 201836 with respect to Nocturnal Polyuria index (NPi).
- To benchmark the efficacy of FE 201836 in relation to desmopressin.

2.2 Endpoints

Primary Endpoint

- Change from baseline in number of nocturnal voids during 12 weeks of treatment.

Secondary Endpoints

- Change from baseline in number of nocturnal voids at Weeks 1, 4, 8 and 12.
- Responder rate defined as 50% reduction in nocturnal voids from baseline at Weeks 1, 4, 8 and 12 during 12 weeks of treatment.
- Change from baseline in NI Diary Total Score at Weeks 1, 4, 8 and 12 and during 12 weeks of treatment.
- Percentage of nights during the treatment period with at most one nocturnal void.

- Percentage of nights during the treatment period with complete response, i.e., with no nocturnal voids.
- Change from baseline in NI Diary Overall Impact Score at Weeks 1, 4, 8 and 12 and during 12 weeks of treatment.
- Patient Global Impression of Improvement (PGI-I) urinary symptoms scores at Weeks 1, 4, 8 and 12.
- Change from baseline in Patient Global Impression of Severity (PGI-S) scores at Weeks 1, 4, 8 and 12.
- Change from baseline in Bother as measured by the Hsu 5-point Likert Bother scale (Hsu, 2015) at Weeks 1, 4, 8 and 12.
- Change from baseline in Insomnia Severity Index (ISI) at Weeks 4, 8 and 12.
- Change from baseline in FUSP at 1, 4, 8 and 12 weeks of treatment and during 12 weeks.
- Change from baseline in nocturnal diuresis rate (hourly) profiles at Week 1 and Week 12.
- Change from baseline in Nocturnal Urine Volume (NUV) at Week 1 and Week 12.

Safety Endpoints

- Incidence and severity of adverse events.
- Incidence of hyponatraemia as measured by serum sodium level throughout the trial.
- Change from baseline in mean 24-hour urine volume at Week 1 and Week 12.
- Clinically significant changes in vital signs and laboratory values.

Exploratory Endpoints

- Responder rates defined as 33%, 60%, 70%, 80%, 90% and 100% reduction from baseline in number of nocturnal voids at Weeks 1, 4, 8 and 12.
- Responder rates defined as 1, 2, and 3 voids reduction from baseline in number of nocturnal voids at Weeks 1, 4, 8 and 12.
- The correlation of copeptin levels to NPi at Visit 2.
- Change from baseline in copeptin levels at Week 12.
- Change from baseline in total sleep time per night at Weeks 1, 4, 8 and 12 and during 12 weeks of treatment.
- Correlation between change from baseline in total sleep time and FUSP at Weeks 1, 4, 8 and 12.
- Change from baseline in NPi at Week 1 and Week 12.

3 Trial design

3.1 Design Considerations

3.1.1 General Considerations

DAWN is a randomized, double-blind, placebo-controlled, response-adaptive, dose-finding trial with Desmopressin Orally Disintegrating Tablet (ODT) as a benchmark. The trial has a screening period with a 2-week life-style intervention period and an enrichment phase with a 1-week single-blind treatment with the highest dose of FE 201836 and a 1-week washout period, followed by a randomization to double-blind, double-placebo treatment with placebo, different doses of FE 201836, or desmopressin.

The trial is designed with two stages:
In the first stage of the trial, 125 subjects are randomized in a 2:2:1 allocation ratio to placebo, 500 µg FE 201836, or desmopressin, and an interim analysis for futility is performed.

In the second stage, the subjects are randomized with the allocation ratio 2:7:1 for placebo, active treatment with FE 201836, or desmopressin, respectively. Subjects allocated to FE 201836 are randomized to 50 µg FE 201836, 100 µg FE 201836, 150 µg FE 201836, 250 µg FE 201836, 350 µg FE 201836, or 500 µg FE 201836 according to allocation probabilities determined in the interim analyses according to pre-specified rules. Subsequent interim analyses will be performed every 8 weeks, where futility analyses will be performed and the randomization allocation probabilities to the different FE 201836 doses will be adjusted.

Based on the Bayesian analysis of a baseline-adjusted dose-response curve, the minimum dose most likely to have more than 85% of the effect of the maximum dose (ED_{85}), and the dose most likely to be the minimally effective dose (MED, defined as the minimum dose that has more than 0.3 voids reduction in nocturnal voids as compared to placebo) will be identified during the reporting of the trial. There are four posterior probabilities of interest as presented in Table 1:

Table 1 Posterior Probabilities of Interest

Quantity of Interest	Posterior Probability within the Trial that:	Objective
$\Pr(\mu_{500} < \mu_{Plb})$	the effect of 500 µg FE 201836 during the 12 weeks of randomized treatment is superior to placebo	Proof-of-Concept
$\Pr(\Delta_{500} < -0.3)$	500 µg FE 201836 has more than 0.3 voids reduction in nocturnal voids as compared to placebo	Futility
Pr(dose d is MED)	dose d is the MED	Establish the MED
Pr(dose d is ED_{85})	dose d is the ED_{85}	Establish ED_{85}

In the trial updates of the allocation probabilities, the two posterior probabilities: Pr(dose *d* is MED) and Pr(dose *d* is ED_{85}), are used in the updated allocation probability for dose *d*. The trial updates are performed as analyses without baseline adjustment, thus, they are not based on the model used for the primary analysis in the reporting of the trial. For additional information, see Section 12.

3.1.2 Sample Size

A maximum of 300 subjects (males and females) will be randomized into the trial. The sample size was based on computer simulations under five sigmoidal dose-response profiles without baseline adjustment. The operating characteristics and associated power were estimated for the different dose-response scenarios by computer simulations using the Fixed and Adaptive Clinical Trial Simulator (FACTS) version 5.6.

The average number of subjects and probabilities for futility and success are presented in Table 2.

Table 2 Operating Characteristics

Dose-Response Scenario	Maximum response (α_2)	ED_{50} (α_3)	Average Number of Subjects	Futility Stopping at First Interim	Total Futility stopping	**Power** (Total Success)
Null	0	0 µg	200	75.8%	97.0%	**3.0%**
I	-0.3	150 µg	277	15.7%	47.9%	52.1%
II	-0.4	200 µg	288	7.9%	25.1%	74.9%
III	-0.5	200 µg	296	2.8%	9.6%	**90.4%**
IV	-0.5	350 µg	295	3.4%	15.4%	**84.6%**
V	-0.6	300 µg	298	1.4%	2.7%	97.3%

For each dose-response scenario, 1,000 trials were simulated in order to estimate the operating characteristics presented in Table 2. The simulations under the null-hypothesis (no effect) demonstrate that the probability of making a false positive conclusion is 3.0%, while under the assumption of a maximum reduction versus placebo of 0.5 voids, the power (probability of success) reaches approximately 85 to 90%.

3.1.3 Trial Stages

In the first of the two trial stages, an interim Bayesian analysis for futility is performed after 125 subjects have been randomized. The futility stopping rule in the first stage is that a less than 10% posterior probability that 500 µg FE 201836 has >0.3 voids reduction in nocturnal voids as compared to placebo ($\Pr(\mu_{500} - \mu_{Plb} < -0.3) < 10\%$).

The futility stopping rule in the second stage is a less than 5% posterior probability that 500 µg FE 201836 has >0.3 voids reduction in nocturnal voids as compared to placebo ($\Pr(\mu_{500} - \mu_{Plb} < -0.3) < 5\%$).

Bayesian estimates based in the dose-response curve without baseline adjustment will be used to update the allocation probabilities for the different doses of FE 201836 (see Section 12).

3.1.4 Trial Epochs

During the screening period, the trial has a 2-week lifestyle-changes period followed by an enrichment phase consisting of a 1-week single-blind active run-in period and a 1-week washout period. The screening period is followed by a 12-week randomized, double-blind treatment period. Thus, the trial has 4 epochs, a lifestyle-changes epoch, a single-blind, active run-in epoch, a washout epoch, and a randomized, double-blind treatment epoch.

3.1.5 Visits and Visit Windows

The scheduled measurement time points are presented in Table 3 for primary, secondary, safety, and exploratory endpoints in the order they occur in Section 2.2.

Table 3 Schedule for Assessments of Endpoints According to the Flow Chart in the Protocol.

Capital X Shows Endpoint-Analyses Time Points, Lowercase x Shows Trial Assessments Not Considered as Endpoints.

Trial Epoch	**Life-style Changes**	**(Start of) Active Run-in**	**(Start of) Washout**	**Randomized Trial Period**				
Visit	**V1**	**V2**	**V3**	**V4 (baseline-RT)**	**V5**	**V6**	**V7**	**V8**
Week					**1**	**4**	**8**	**12**
Planned Day		**−14**	**−7**	**1**	**8±2**	**28±3**	**56±3**	**84±7**
Endpoints								
Primary Endpoint								
Number of nocturnal voids		x	x	x	X	X	X	X
Secondary Endpoints								
Responder rates derived from number of nocturnal voids					X	X	X	X
NI Diary Total Score		x		x	X	X	X	X
NI Diary Overall Impact Score		x		x	X	X	X	X
PGI-I					X	X	X	X
PGI-S		x		x	X	X	X	X
Bother				x	X	X	X	X
ISI		x		x		X	X	X
FUSP		x	x	x	X	X	X	X
Nocturnal diuresis rate (=NUV / Total sleep)		x	x	x	X			X
NUV		x	x	x	X			X
Safety Endpoints								
Adverse events [a]		X	X	X	X	X	X	X
Hyponatremia serum sodium level	X		X	[b]	X	X	X	X
Total daily urine volume		x		x	X			X
Vital signs	x	X	X	X	X	X	X	X
Laboratory values: Clinical chemistry, haematology, urinalysis	x			X	X	X	X	X
Laboratory values: Coagulation factors	x			X		X		X

Trial Epoch	**Life-style Changes**	**(Start of) Active Run-in**	**(Start of) Washout**	**Randomized Trial Period**				
Visit	**V1**	**V2**	**V3**	**V4 (baseline-RT)**	**V5**	**V6**	**V7**	**V8**
Week					**1**	**4**	**8**	**12**
Planned Day		**−14**	**−7**	**1**	**8±2**	**28±3**	**56±3**	**84±7**
Endpoints								
Exploratory Endpoints								
NPi (=NUV / Total daily urine volume)		X		x	X			X
Copeptin		X		x				X
Total sleep		x	x	x	X	X	X	X

[a] AEs are also recorded at a telephone follow-up scheduled at Day −10±1

[b] Serum Sodium is also measured at Visit 4 as a part of the standard laboratory measurments

For post-randomization Voiding Diary endpoints, the date of completion of the diary (Night 3), will be used for the allocation of the data according to the visit windows. For all other endpoints including pre-randomization Voiding Diary data, the corresponding actual site visit day will be used for the allocation of the data according to the visit windows. The analysis day (ADY) is defined as date of corresponding site visit (Night 3 for post-randomization Voiding Diary data) – date of Visit 4 (day of randomization / baseline of the double-blind treatment period). Visit 4 is defined as Analysis Day 0 and has no visit window for the corresponding site visit date.

For the screening period (the lifestyle-changes epoch, the active run-in epoch, and the washout epoch), the visit windows for all endpoints are linked to the start of the active run-in epoch (the baseline for the enrichment period), that is, the date of the scheduled Visit 2 (VISITNUM=2): All visits before the date of VISITNUM=2 are allocated to the Visit 1 window. All visits after the date of VISITNUM=2 and before the date of VISITNUM=4 are allocated to the Visit 3 window.

Prioritization between multiple visits with non-missing assessment of the endpoint within the same visit window for the screening period will be performed according the following rules, unless otherwise specified:

1. If the assessment is non-missing at the scheduled visit, the assessment at the scheduled visit is used.
2. If the assessment is missing at the scheduled visit, the non-missing assessment closest to the scheduled visit is used.
3. If the assessment is missing at the scheduled visit and the two closest, non-missing assessments from unscheduled visits in the same visit window have equal distances from the date of the scheduled visit, then the latest one will be selected (only relevant for Visit 3).

For the randomized, double-blind treatment epoch, different visit windows will be used for different endpoints. The endpoints fall mainly into two major categories according to when the endpoint is assessed.

Category I comprises efficacy and safety endpoints assessed at all visits during the randomized-treatment period. Efficacy endpoints are Number of Nocturnal Voids and Responder outcomes derived from number of nocturnal voids, NI Diary Total Score, NI Diary Overall Impact Score, PGI-I. PGI-S, Bother, FUSP, and Total Sleep. Safety endpoints are Hyponatremia Serum Sodium Level, Clinical Chemistry, Haematology, Urinalysis, Weight, and Vital Signs.

Category II comprises the efficacy endpoints assessed only at Visits 5 and 8: Nocturnal Diuresis Rate, NUV, and NPi; and the safety endpoint: Total Daily Urine Volume.

The remaining endpoints, the efficacy endpoints ISI and Copeptin and the safety endpoint Coagulation factors, are assessed at different visits. All visit windows for the randomized-treatment period are specified below (Table 4):

Table 4 Visit Windows Using Analysis Day Calculated as Site Visit Date – Date of Visit 4.

	Visit windows [a]			
Visit Number	**5**	**6**	**7**	**8**
Visit Week	**Week 1**	**Week 4**	**Week 8**	**Week 12**
Planned Day According to Protocol	**8**	**28**	**56**	**84**
Target day [b]	7	28	56	84
Category I Endpoints	[1, 17]	[18, 42]	[43, 70]	[71, ∞[
Category II Endpoints	[1, 45]	NA	NA	[46, ∞[
ISI	NA	[1, 42]	[43, 70]	[71, ∞[
Copeptin	NA	NA	NA	[1, ∞[
Coagulation factors	NA	[1, 55]	NA	[56, ∞[

[a] using analysis day derived from the corresponding site visit date

[b] to be used when prioritizing multiple assessments of an endpoint within the visit window

Prioritization between multiple visits with non-missing assessment of the endpoint within the same visit window will be performed according the following rules, unless otherwise specified:

1. If more than one non-missing assessment of an endpoint fall within the same visit window, the one closest to the target day will be used for that endpoint.
2. If two non-missing assessments in the same visit window have equal distances from the target day, the latest one will be used.

For the first safety interim analysis, the actual visit (VISITNUM) will be used. For the futility interim analyses, the site visit date and the visit windows specified in Table 4 will be used.

4 Major Protocol Deviations

The following protocol deviations are considered major in the sense that they lead to exclusion of subject or subject data from the Per Protocol Set (see Section 5.2.2):

- The subject is excluded in case of <2 nocturnal voids (average over 3 diary nights) prior to Visit 2 or Visit 4
- The subject is excluded in case of NPi ≤33% prior to Visit 2
- The subject is excluded in case of <20% decrease in nocturnal diuresis rate (mL/min) from during active run-in (Visit 2 to Visit 3)
- The subject is excluded in case of maximum voided volume <200 mL prior to Visit 2
- The subject is excluded in case of actual treatment not in accordance with randomized treatment
- Visit-specific data are excluded in case of significant non-compliant treatment administration for the visit. For each visit, this is defined as e-Diary-registered intake of trial medication for less than half of the nights for which number of nocturnal voids could be derived
- The subject is excluded in case of significant non-compliant e-Diary completion at Visit 4 defined as the Voiding Diary filled out more than 1 week before the site visit for Visit 4
- For Visits◦6, 7 and 8, visit-specific data are excluded in case of significant non-compliant e-Diary completion at the visit defined as the Voiding Diary completed more than 2 weeks before or after the planned day according to the protocol (28, 56, and 84 days after Visit 4, respectively). For Visit 5, significant non-compliance is defined as the Voiding Diary completed more than 14 days after Visit 4
- Visit-specific data are excluded in case of current treatment with any of the following prohibited medications:
  - Thiazide diuretics [ATC code: C03AA]
  - V2-receptor antagonists/agonists (e.g., vaptans/desmopressin, vasopressin) [ATC code: C03XA/H01BA]
  - Loperamide [ATC code: A07DA03]
  - Botulinum toxin (cosmetic non-urological use is acceptable) [ATC code: J06AA04]
  - Valproate [ATC code: N03AG01]

Serious unforeseen deviations deemed to impact the primary endpoint of the trial may additionally be rated as major protocol deviations by the sponsor on the basis of a blinded review of data before declaration of clean-file and lock of database. The list of major protocol deviations will be detailed and documented in the clean file document prior to database release to the trial statistician.

5 Analysis sets

5.1 Enrichment-Period Analysis Sets

5.1.1 Intention-To-Treat Analysis Set for the Enrichment Period (ITT-ENR)

The Intention-to-Treat analysis set for the enrichment period (ITT-ENR) comprises all subjects found to be eligible to enter the enrichment period (comprising 1 week of active run-in and 1 week of washout) at Visit 2. All ITT-ENR analyses will be according to the planned treatment allocation.

5.1.2 Safety Analysis Set for the Enrichment Period (Safety-ENR)

The safety analysis set for the enrichment period (Safety-ENR) comprises all subjects who received investigational medicinal product (IMP) at Visit 2 and did not return all IMP unused at or before Visit 4, based on the drug accountability. This also includes any subject who was found not eligible at Visit 2 but received at least one dose of IMP at or after Visit 2 and before Visit 4 and did not return all IMP unused at or before Visit 4. All safety analyses will be according to actual treatment received.

5.2 Randomized-Treatment-Period Analysis Sets

5.2.1 Intention-To-Treat Analysis Set for the Randomized-Treatment Period (ITT-RT)

The Intention-to-Treat analysis set for the randomized-treatment period (ITT-RT) comprises all subjects randomized at Visit 4.

5.2.2 Per Protocol (PP) Analysis Set for the Randomized-Treatment Period

The Per Protocol analysis set comprises all subject data (that is, all combinations of subject and visit) in the ITT-RT analysis set excluding subject data that meet any of the major protocol deviations as defined in Section 4. Subjects for whom all post-baseline data are excluded are not included in the Per Protocol Set.

5.2.3 Safety Analysis Set for the Randomized-Treatment Period (Safety-RT)

The safety analysis set for the randomized-treatment period (Safety-RT) comprises all subjects who received IMP at Visit 4 and did not return all IMP unused, based on the drug accountability, and have had at least one safety assessment after randomization (Visit 4). This also includes any subject who was found not eligible at Visit 4 but who received at least one dose of IMP at or after Visit 4 and did not return all IMP unused. All safety analyses will be according to actual treatment received.

6 Trial population

6.1 Demographics and Other Baseline Characteristics

6.1.1 General Considerations

The trial has two baselines, one for the enrichment period (denoted baseline-ENR) including the latest data obtained before or at Visit 2, and one for the randomized-treatment period (denoted baseline-RT) including the latest data obtained before or at Visit 4 unless otherwise specified.

6.1.2 Demographics

Baseline-ENR values of the demographics variables: age, sex, race, and ethnic origin, height, weight and BMI will be summarized overall for the ITT-ENR and the Safety-ENR analysis sets. Baseline-RT demographics will be summarized overall and by treatment group for the ITT-RT analysis set.

6.1.3 Vital Signs at Baseline

Baseline-ENR vital signs will be summarized overall for the ITT-ENR analysis set. Baseline-RT vital signs will be summarized overall and by treatment group for the ITT-RT analysis set.

6.1.4 Physical Examination at Baseline

Physical examination at baseline-ENR will be summarized overall for the ITT-ENR analysis set, and physical examination at baseline-RT will be summarized overall and by treatment group for the ITT-RT analysis set.

6.1.5 Laboratory Efficacy Parameters at Baseline

Baseline-ENR copeptin will be summarized for the ITT-ENR analysis set. Baseline-RT copeptin will be summarized overall and by treatment group for the ITT-RT analysis set.

6.1.6 e-Diary Assessments at Baseline

The baseline registrations in the e-Diary include a Voiding Diary, a Nocturia Impact Diary, PGI-S, ISI, and Bother. Based on the baseline registrations in the Voiding Diary, baseline values for several variables are derived: number of nocturnal voids, nocturnal urine volume, NPi, nocturnal diuresis rate, FUSP, and total sleep. The Nocturia Impact Diary is used to derive the baseline values of the two NI Diary endpoints, NI Diary Total Score and the overall QoL impact question.

Baseline-ENR values of the e-Diary variables will be summarized overall for the ITT-ENR analysis set. Baseline-RT values of the e-Diary variables will be summarized overall and by treatment group for the ITT-RT analysis set.

6.2 Medical History

Medical history recorded at the screening visit will be coded using the latest version of MedDRA and tabulated by System Organ Class (SOC) and preferred term (PT) and randomized-treatment group (including non-randomized subjects as a separate treatment group) for the Safety-ENR analysis set.

7 Subject Disposition

Presentations of subject dispositions will be produced for the overall trial population and by sex.

All subjects screened will be accounted for. All screening failures, protocol violations, and discontinuations from trial will be summarized with frequency and percentage for each category of screening failure reason, protocol violation, and reason for discontinuation, respectively.

All screening failures will be listed in a data listing including the reason(s) for screening failure.

Subject disposition for the ITT-RT analysis set (Section 5.2.1) will be tabulated by treatment group. Reasons for not being randomized at Visit 4 will be categorized and summarized for the ITT-ENR (Section 5.1.1) analysis set. Subject disposition for the safety analysis sets (Sections 5.1.2 and 5.2.3), will be presented separately.

For the ITT-RT analysis set, the treatment-arm (placebo, FE 201836 pooled, desmopressin pooled) differences in time to early discontinuation will be tested using the Log-Rank test, and, for each treatment arm separately, the one-minus-Kaplan–Meier (1-KM) plots of time to early discontinuation will be displayed differentiated by reason for discontinuation. In these analyses, completers will be censored at the maximum of 84 days and the latest observed day of early discontinuation. All early discontinuations will be summarized by time of and reason for discontinuation.

8 Prior and Concomitant Medication

Prior and concomitant medication will be summarized by the Anatomical Therapeutic Chemical Classification System (ATC) classification 1st level (alphabetically), ATC classification 2nd level (in decreasing order of frequency) and treatment group. These medications will be tabulated for the Safety-ENR analysis set according to

1) Prior medication taken exclusively prior to any treatment with IMP, i.e., with stop date before the date of dispense of first single-blind IMP;
2) Concomitant medication taken during the enrichment period, i.e., medication that was not stopped before the date of first single-blind IMP administration and not started after the date of first double-blind IMP administration;
3) Concomitant medication taken during the randomized-treatment period, i.e., medication that was not stopped before the date of first double-blind IMP administration and not started after the End-of-Trial Visit.

If the timing of the dose of a concomitant medication cannot be established in relation to the administration of randomized IMP, it will be considered as concomitant medication in both the enrichment period and the randomized-treatment period.

9 Exposure and Treatment Compliance

9.1.1 Extent of Exposure

9.1.1.1 General Considerations

There are two independent sources for assessment of the extent of exposure: investigator's registration of amount of IMP dispensed and amount of IMP returned, and the subject's e-Diary registrations.

The subjects are always dispensed two sets of IMP, one set of bottles and one set of ODT. Therefore, there are two sets of registrations of IMP accountability at each visit, one for the number of bottles dispensed and returned, and one for the number of ODT dispensed and returned. The investigator registers the amount of IMP dispensed in number of bottles and number of ODT at the scheduled Visits 2, 4, 5, 6, and 7 with dispense of IMP; and the amount of IMP returned in number of bottles and number of ODT at the scheduled Visits 3, 5, 6, 7, and End-of-Trial visit with accountability of IMP.

The subject registers whether he/she took the trial medication in the e-Diary on a daily basis between Visits 2 and 3, and from Visit 4 to End-of-Trial Visit.

The two sources will be summarized separately, and in addition, the agreement between the two sources will be quantified and summarized.

9.1.1.2 Extent of Exposure Based on the Investigator-Assessed Amount of IMP Dispensed and Returned

Extent of exposure is measured both as duration of treatment in number of days and as number of doses. The exposure in number of bottles and the exposure in number of ODT will be summarized separately.

Duration for the active run-in period is derived as the date of the earliest of Visit 3 and End-of-Trial Visit minus the date of Visit 2. Number of doses for the active run-in period is derived as number of doses dispensed at Visit 2 minus number of doses returned at the earliest of Visit 3 and End-of-Trial Visit. Duration and number of doses during the active run-in period will be summarized overall and by sex for the ITT-ENR analysis set and for the Safety-ENR analysis set.

For each of the post-randomization visits with registration of amount of IMP returned, the duration in the preceding period between visits will be derived as the difference between the date of the current visit and the date of the previous visit with IMP accountability. The number of doses will be calculated as the number of doses dispensed at the previous visit with IMP accountability minus the number of doses returned at the current visit. If two or more visits with IMP accountability fall into the same visit window, the duration will be summed across the visits and the number of doses will be summed across the visits.

The duration and the number of doses will be summarized by treatment group and visit, overall and by sex, for the ITT-RT analysis set and for the Safety-RT analysis set.

Duration for the whole randomized-treatment period is derived as the date of the End-of-Trial Visit minus the date of Visit 4 and will be summarized by treatment group for the ITT-RT analysis set and for the Safety-RT analysis set. Number of doses for the whole randomized-treatment period is derived as the sum of the doses across the post-randomization visits and will be summarized by treatment group for the ITT-RT analysis set and for the Safety-RT analysis set.

Total duration will be derived as the difference between the date of the End-of-Trial Visit and the date of Visit 2. Total extent of exposure in number of doses will be derived as the sum of the doses in the enrichment period and the doses in the randomized-treatment period. Total duration and total extent of exposure in number of doses will be summarized by treatment group for the ITT-RT analysis set and for the Safety-RT analysis set.

9.1.1.3 Extent of Exposure Based on the Subject's e-Diary Registration

When deriving extent of exposure based on the e-Diary, days with Voiding Diary assessments as well as days without Voiding Diary assessments will be considered.

For each visit with an e-Diary assessment of IMP intake, the duration of exposure in days is derived as 1 + the difference between the date of the latest intake of IMP after the previous visit and the date of the earliest intake of IMP after the previous visit, and the extent of exposure in number of doses will be derived as the number of days where the subject answered Yes to "*Did you take your study medication [during this diary day]?*". In case two or more visits with an e-Diary assessment of IMP intake fall into the same visit window, the duration will be derived as the difference between the latest date of intake of IMP in any of the diaries and the earliest date of intake of IMP in any of the diaries, and the number of doses will be derived as the sum of doses across the diaries.

Duration and number of doses at Visit 3 will be summarized overall and by sex for the ITT-ENR analysis set and for the Safety-ENR analysis set.

Duration and number of doses will be summarized by treatment group and visit, overall and by sex, for the ITT-RT analysis set and for the Safety-RT analysis set.

9.1.1.4 Agreement of the Two Sources

For each visit with e-Diary assessments of the exposure (scheduled Visits 3, 5, 6, 7, and 8), the number of doses according to the e-Diary in Section 9.1.1.3 will be subtracted from the investigator-assessed number of doses of bottles and ODT, respectively, derived for the same period according to Section 9.1.1.2. The differences will be summarized by treatment group, sex, and visit; Visit 3 will be summarized for the ITT-ENR analysis set and for the Safety-ENR analysis set; the post-randomization visits will be summarized for the ITT-RT analysis set and for the Safety-RT analysis set.

9.1.2 Treatment Compliance

Compliance will be derived separately for each of the three exposure measures of number of doses: the number of bottles and the number of ODT (Section 9.1.1.2), and the number of doses according to the e-Diary (Section 9.1.1.3). For each exposure measure of number of doses, compliance in % during a specific period (that is, a period between two visits, the active run-in period, the randomized-treatment period, or total exposure duration) will be derived as the exposure measure of number of doses for the period divided by the corresponding exposure duration in days for the same period calculated in Section 9.1.1.2 as the end date minus the start date of the period, and multiplied by 100 to obtain a percentage.

Compliance during the active run-in period will be summarized as overall and by sex for the ITT-ENR analysis set and for the Safety-ENR analysis set.

Compliance will be summarized by treatment group and visit, overall and by sex, for the ITT-RT analysis set and for the Safety-RT analysis set.

In addition, for each visit with Voiding Diary assessments, treatment compliance will be derived for the nights where the number of nocturnal voids can be derived. The treatment compliance in % during Voiding Diary assessment will be derived as the number of nights where the number of nocturnal voids can be derived and the subject answered Yes to "*Did you take your study medication during this diary day?*" divided by the total number of nights where the number of nocturnal voids can be derived in the Voiding Diary for the visit, and multiplied by 100 to obtain a percentage. Treatment compliance during Voiding Diary assessment will be summarized as a categorical variable by treatment group and visit for the ITT-RT analysis set and for the PP analysis data set.

10 Efficacy

10.1 General Considerations

For endpoints obtained from the Voiding Diary prior to a visit, the average of the endpoint across the 3 days will be interpreted as the cross-sectional observation of the endpoint in question at the specific visit, with the exception of the two endpoints, "Percentage of nights during the treatment period with at most one nocturnal void" and "Percentage of nights during the treatment period with complete response, i.e., with no nocturnal voids".

If the baseline number of nocturnal voids is missing for a reason that can be considered to be missing completely at random (e.g., missing due to technical problems of the eDiary), the subject will not included in any analyses that involves baseline measurements of nocturnal voids or other variables derived using baseline measurements of nocturnal voids.

The Bayesian dose-response analyses are conducted on aggregated averages, that is, outcomes averaged across preceding post-randomization visits. The posterior mean, median, and 95% credibility interval (2.5 and 97.5 percentiles of the posterior distribution) of quantities of interest as well as posterior probabilities of interest will be reported.

Frequentist inferences will rely on the appropriate choice of model and report p-values corresponding to the statistical test of the hypothesis of "no difference / no association" against the alternative of "any difference / any (linear) association". The level of significance is set at 5% (two-sided).

Treatment comparison refers to the pair-wise comparison between each of the different doses of active treatment with FE 201836 and placebo. There are no planned adjustments for multiplicity when pair-wise testing the active treatment with different doses of FE 201836 versus placebo. Desmopressin is included as a benchmark treatment only. None of the primary, secondary or explorative analyses will include data from subjects treated with desmopressin. Explorative benchmarking analyses are further described in Section 10.5.

10.2 Primary Efficacy Endpoint

10.2.1 Primary Analysis

The primary endpoint is the aggregated average of change from baseline-RT in number of nocturnal voids after 12 weeks of treatment, that is, averaged across all post-randomization visits. The objectives of the primary analysis are to establish Proof-of-Concept of FE 201836 and to estimate the dose-response profile of FE 201836. The analyses will be performed on the ITT-RT analysis set.

The overall trial outcome is proof-of-concept evaluated by the posterior probability that 500 µg FE 201836 has superior efficacy as compared to placebo ($Pr(\mu_{500}<\mu_{Plb})>97.5\%$) based on a Bayesian analysis of the dose-response profile at Week 12 for the ITT-RT analysis set not including the active comparator desmopressin.

The dose-response profile will be characterized by the posterior estimates of the mean dose-response at each dose, μ_d, corresponding to two nocturnal voids at baseline and the mean dose-response contrast versus placebo, Δ_d, with their associated 95% credibility intervals. The posterior probabilities $Pr(\Delta_d<-0.3)$, $Pr($dose d is MED$)$, and $Pr($dose d is $ED_{85})$, will be tabulated. The dose-response curve relative to placebo will also be presented graphically with its 95% credible interval in the plots illustrating the therapeutic window (Section 10.4). No multiplicity adjustment will be performed.

The dose-response profiles are determined using the following model: Let $x_{t,i}$ denote the observed cross-sectional change from baseline-RT in nocturnal voids (averaged across the nights in the Voiding diary corresponding to the visit) and let $y_{t,i}$ denote the aggregated average change from baseline for subject i, averaged across post-randomization visits up to time t (for visit windows see Section 3.1.5):

$$y_{t,i} = \frac{\sum_{k=1}^{t} x_{k,i}}{t},\ t = 1,2,3,4,$$

where t denotes the nominal post-randomization time points reflecting Visits 5, 6, 7 and 8, corresponding to Weeks 1, 4, 8, and 12, respectively. The final endpoint, $y_{4,i}$ is modelled as being an observation of

$$Y_{4,i} \sim N(\beta\cdot(x_{0,i}-2) + \mu_d, \sigma^2), \quad (1)$$

where $x_{0,i}$ is the baseline-RT number of nocturnal voids, σ^2 is the error variance, and μ_d describes the dose-response relationship. μ_d is assumed to follow a four-parameter sigmoidal model:

$$\boldsymbol{\mu_d = \alpha_1 + \frac{(\alpha_2-\alpha_1)d^{\alpha_4}}{d^{\alpha_4}+\alpha_3{}^{\alpha_4}}}, \quad (2)$$

where d is the dose measured in 100 µg (so d=0, 0.5, 1, 1.5, 2.5, 3.5, or 5), and the interpretation of the parameters are: α_1 is the placebo response for a subject with 2 nocturnal voids at baseline; α_2 is the maximum response for a subject with 2 nocturnal voids at baseline; α_3 is the dose (ED_{50}) at which 50% of the maximum effect is obtained; and α_4 is the slope parameter. In the primary analysis of $y_{4,i}$, the model parameters $\theta = (\beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \sigma^2)$ are modelled with the prior distributions $\beta\sim N(0, 10^2)$; $\alpha_1\sim N(-1, 3^2)$; $\alpha_2\sim N(-0.5, 2^2)$; $\alpha_3\sim N^+(2, 2^2)$; $\alpha_4\sim N^+(5, 3.5^2)$; and $\sigma^2\sim IG(0.5, 0.605)$, where the notation $N^+(\eta, \zeta^2)$ refers to a positive truncated normal distribution, and the $IG(a, b)$ is the inverse gamma distribution.

Missing data due to drop out or insufficient Voiding Diary data at Visit 8 will be imputed using a longitudinal model applied to the observed data including all visits, corresponding to a missing-at-random assumption. The longitudinal model of $y_{t,i}$ is modelled conditional on an individual parameter, δ_i, modelling $y_{t,i}$ as

$$\boldsymbol{y_{t,i} = e^{\xi_t}\left(\beta(x_{0,i} - 2) + \mu_d + \delta_i\right) + \varepsilon_{t,i}} \quad \textbf{(3)}$$

where ξ_4=0. Bayesian analysis will be based on the same priors as in the primary analysis described above for the β and the α parameters, while the remaining parameters are modelled with the priors: $\xi_t \sim N(0, 1)$; $\delta_i \sim N(0, \tau^2)$ with $\tau^2 \sim IG(0.5, 0.36125)$, and $\varepsilon_{t,i} \sim N(0, \lambda_t^2)$ with $\lambda_t^2 \sim IG(0.5, 0.125)$.

A total of 1000 imputations of the missing $y_{4,i}$ will be performed. In the multiple imputation of the missing $y_{4,i}$ for the primary analysis, a sample of size 1000 of the model parameters, $\pi = (\xi_1, \xi_2, \xi_3, \beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \delta_i\ (i=1,...,300), \lambda_1^2, \lambda_2^2, \lambda_3^2, \lambda_4^2)$, is first sampled from the posterior for the Bayesian analysis of the longitudinal model, using the seed=2720, a burn-in of 10 000 and a thinning factor of 50. Then for each set of model parameters, π_s, the imputed value of a missing $y_{4,i}$ is obtained by adding a normally distributed error, $\varepsilon_{4,i,s}$, with seed=152 to the conditional mean of $y_{4,i}$ (conditional on $\delta_{i,s}$), giving a single imputation, $y_{4,i,s}$, thereby capturing all the variability in each imputed $y_{4,i,s}$.

For each of the imputed data sets, the Bayesian analysis of the (possibly imputed) $y_{4,i}$ will be performed with a seed=10000+s using a burn-in of 10 000, and a single set of the model parameters θ_s is obtained as number $50 \cdot s$ after the burn-in of the generated posterior values of θ. The set $\{\theta_s; s=1,...,1000\}$ will be used to derive the μ_d and Δ_d with 95% credibility intervals and the posterior probabilities, $Pr(\Delta_d < -0.3)$, $Pr($dose d is MED$)$, and $Pr($dose d is $ED_{85})$.

10.2.2 Sensitivity Analyses

The following sensitivity analyses will be conducted:

1. The mean, the median, and the 95% credibility intervals for μ_d, Δ_d, $\exp(\xi_1)$, $\exp(\xi_2)$, and $\exp(\xi_3)$; as well as the posterior probabilities, $Pr(\Delta_d < -0.3)$, $Pr($dose d is MED$)$, and $Pr($dose d is $ED_{85})$, will be presented for the longitudinal analysis used in the imputation of $y_{4,i}$.
2. The primary-analysis approach according to the planned treatment allocation will be carried out on the PP analysis set.
3. The primary-analysis approach according to the planned treatment allocation will be carried out on the ITT-RT analysis set, replacing μ_d by α_1 in the conditional mean of missing $y_{4,i}$ when imputing from the longitudinal model, that is, using a Missing Not At Random (MNAR) placebo model for the imputation of subjects who drop out.
4. The primary analysis according to the planned treatment allocation will be carried out on the ITT-RT analysis set using priors with twice as large scale parameters (standard deviations for normal distributions) than used for the primary analysis.
5. A Bayesian analysis will be performed for the observed value at Week 12 for completers, using a seed=10 000 and the same burn-in and thinning factor as specified for the longitudinal analysis, obtaining a final posterior sample size equal to the posterior sample size obtained in the primary analysis.

6. The primary-analysis approach but using the actual treatment allocation will be carried out on the Safety-RT analysis sets.
7. A mixed-effect model for repeated measurements (MMRM) analysis according to the planned treatment allocation will be performed on the ITT-RT analysis set using Markov chain Monte Carlo (MCMC) to impute intermediate missing 1000 times with seed=742738, and an MNAR placebo-based pattern mixture model (see Ratitch 2011) to impute after drop out for each of the 1000 imputed data set with monotone missing pattern, using the seeds 841621, 745274, 830367, and 995004 for the imputation of missing at Week 1, Week 4, Week 8, and Week 12, respectively. The systematic effects will comprise visit number and treatment-by-visit as categorical variables, and baseline-RT number of nocturnal voids minus 2 as a linear covariate interacting with visit number. An unstructured covariance matrix will be used. The estimated contrasts to placebo at Week 12 will be presented with 95% confidence intervals. Based on the same sets of imputed data, the log-linear trend across active doses (replacing the categorical treatment variable with an indicator variable for treatment with Placebo and a linear variable, $\log_2(d/0.5)$) and the linear trend across all doses (replacing the categorical treatment variable with the linear variable, d) will be estimated with 95% confidence limits.

10.2.3 Descriptive analysis

The mean aggregated average and mean aggregated average change from baseline for each dose at each visit will be tabulated together with 95% confidence intervals as a part of the explorative benchmarking analyses described in Section 10.5.

10.3 Secondary Efficacy and Exploratory Analyses

10.3.1 General Considerations

The following main categories of inferential analyses will be performed for the secondary efficacy endpoints:

- Bayesian estimation of the sigmoidal dose-response curve will be used for aggregated averages and for the aggregated responder variable, 50% reduction from baseline-RT in number of nocturnal voids, defined as the aggregated average change from baseline being less or equal to 50% of the baseline-RT number of nocturnal voids. For details, see Section 10.3.3.
- MMRM will be used to evaluate treatment effects on cross-sectional change from baseline for quantitative variables. For details, see Section 10.3.4.
- General Estimating Equations with the logit as link function will be used for cross-sectional binary responder variables. For details, see Section 10.3.5.
- Analysis of covariance (ANCOVA) will be used for the single-time-point quantitative variables: percentage of nights during treatment period with at most 1 nocturnal void and with a complete response, respectively, and change from baseline in copeptin levels. For details, see Section 10.3.6.

- Correlation between endpoints will be reported using Pearson correlation coefficients and linear regression coefficients, regressing total sleep on FUSP and NPi on copeptin.

Table 5 presents the planned analyses for all endpoints. All analyses will be performed on the observed data in the ITT-RT analysis set unless otherwise specified. Visit windows for each endpoint will be as described in Section 3.1.5.

Table 5 Analysis of Endpoints

Endpoint	Analysis category				
	Bayesian Estimation of Dose-response Curve for Aggregated Averages	**MMRM**	**Repeated logistic regression (Generalized Estimating Equation [GEE])**	**ANCOVA**	**Correlation Analysis**
Efficacy Endpoints					
Aggregated average change from baseline during the 12 weeks randomized-treatment period for number of nocturnal voids	Primary	Sensitivity, Therapeutic Window			
Change from baseline in number of nocturnal voids at the visit [a]		Secondary			
Rate of response defined as 50% reduction in nocturnal voids from baseline based on aggregated average change from baseline	Secondary		Secondary		
Rate of response defined as 50% reduction in nocturnal voids from baseline [a]			Secondary		
Rate of response defined as 33%, 60%, 70%, 80%, 90% and 100% reduction from baseline in number of nocturnal voids [a]			Explorative		
Rate of response defined as 1, 2, and 3 voids reduction from baseline in number of nocturnal voids [a]			Explorative		
Percentage of nights during treatment period with at most one nocturnal void [b]				Secondary	
Percentage of nights during treatment period with complete response [b]				Secondary	
Change from baseline in NUV [c]	Secondary	Secondary			
Change from baseline in FUSP [a]	Secondary	Secondary			Explorative [d]
Change from baseline in total sleep time per night [a]		Explorative			
Change from baseline in nocturnal diuresis rate (hourly) [c]		Secondary			
Change from baseline in NPi [c]		Explorative			
Change from baseline in NI Diary Total Score [a]	Secondary	Secondary			
Change from baseline in NI Diary Overall Impact Score [a]	Secondary	Secondary			

Endpoint	Analysis category				
	Bayesian Estimation of Dose-response Curve for Aggregated Averages	**MMRM**	**Repeated logistic regression (Generalized Estimating Equation [GEE])**	**ANCOVA**	**Correlation Analysis**
PGI-I urinary symptoms score [a, e]		Secondary			
Change from baseline in PGI-S score [a]		Secondary			
Change from baseline in Bother as measured by the Hsu 5-point Likert Bother Scale [a]		Secondary			
Change from baseline in ISI [f]		Secondary			
Copeptin levels at Visit 2 [g]					Explorative
Change from baseline in copeptin levels [h]				Explorative	
Safety Endpoints					
Change from baseline in mean 24-hour urine volume [c]		Safety			
Lowest observed change from baseline-RT in Serum Sodium				Alternative safety indicator for therapeutic window	
Incidence and severity of Adverse Events	Section 11.3.1				
Adverse Event of Special Interest: Hyponatraemia Episodes	Section 11.2.1				
Clinically significant changes in laboratory values and vital signs	Sections 11.3.2 and 11.3.3				

[a] Endpoint assessed at Weeks 1, 4, 8 and 12.

[b] Endpoints are derived using all diary data collected during the treatment period.

[c] Endpoint assessed at Weeks 1 and 12.

[d] Correlation between changes in FUSP and total sleep time at Weeks 1, 4, 8 and 12.

[e] This questionnaire is retrospective and raw scores will be used in the analysis (not changes from baseline) and the ANCOVA model will not be adjusted for baseline value.

[f] Endpoint assessed at Weeks 4, 8 and 12.

[g] Correlation with NPi at Visit 2.

[h] Endpoint assessed at Week 12.

Illustration of the therapeutic window is described separately in Section 10.4 and safety related endpoints are further described in Section 11.

Explorative analyses by sex are planned for all endpoints, including correlation analyses. For these the following are noted:

- Estimates of the sex-specific dose-response for the aggregated averages after 12 weeks of treatment will follow the Bayesian analysis described in Sections 10.2.1 and 10.3.3 by conducting a separate analysis for each sex. Possible differences between the sex-specific models will be investigated.
- The regression models will be supplemented with an analysis considering sex as well as all interactions between sex and the effects specified in the all-subjects analysis, including any interaction terms specified in the all-subject analysis. Possible differences between the treatment effects for men and women will be evaluated by the treatment-by-visit-by-sex (treatment-by-sex for single-visit outcomes) interaction term with the effect of treatment-by-visit (treatment for single-visit outcomes) included in the model.

10.3.2 Scoring of Patient-Reported Outcomes

The post-randomization, visit-specific, cross-sectional change from baseline-RT, $x_{t,i}$, in number of nocturnal voids is defined as the average across the nights in the Voiding Diary corresponding to the single visit. Regarding assigning number of voids for a single night, the following rules are used:

1. Voids less than 5 minutes apart will be collapsed into a single void and the corresponding volumes will be summed to a total volume. Voids on different sides of a time threshold (5 minutes after bed time with the intention to sleep at night, rising in the morning, and 30 minutes after rising in the morning) will not be collapsed.
2. If day/night cannot be assigned due to missing go-to-bed or got-up time, number of nocturnal voids and NUV for that 24-hour period is set to missing.
3. If day/night can be assigned and there are no voids during the night period, number of nocturnal voids is set to 0 for that 24-hour period and NUV is set to 0 plus the volume of the first void within 30 minutes of rising in the morning.
4. If day/night can be assigned, and there are a positive number of voids during both the night period, the number of voids during the night period is used, whether or not the volume of every void is registered. If at least one volume is not registered, NUV is set to missing for that 24-hour period.

The post-randomization aggregated average change from baseline-RT, $y_{t,i}$, in number of nocturnal voids is defined as the average of the cross-sectional changes from baseline-RT, $x_{s,i}$, across the post-randomization visits up to and including the current visit. If the cross-sectional change from baseline-RT, $x_{t,i}$, is missing at a single visit, the corresponding aggregated average change from baseline-RT, $y_{t,i}$, will be set to missing.

The aggregated average change from baseline being less or equal to 50% of the baseline-RT number of nocturnal voids, $z_{50\%,t,i}$, is derived from the corresponding aggregated average change from baseline in number of nocturnal voids, $y_{t,i}$, as $z_{50\%,t,i} = 1$ if $y_{t,i}/x_{0,i} \leq -0.5$, and $z_{50\%,t,} = 0$ otherwise.

The visit-specific, cross-sectional responder variables related to number of nocturnal voids (33%, 50%, 60%, 70%, 80%, and 90% reduction from baseline; and 1, 2, and 3 voids reduction from baseline) are derived by categorization of the corresponding $x_{t,i}$ divided by the baseline-RT number of nocturnal voids, $x_{0,i}$.

The percentages of nights with response defined as at most one nocturnal void and complete response (no nocturnal voids), respectively, are derived as the percentage of nights with the specific response among all single, post-randomization nights in the Voiding Diary for which number of nocturnal voids can be derived.

The NUV for a single night is the total urine volume in mL from 5 minutes after bedtime with the intention to sleep including the first void within 30 minutes of rising in the morning. The NUV for a visit is calculated as the average of the single-night values. Aggregated values are derived similarly to the derivation of the aggregated values for number of nocturnal voids.

The FUSP for a single night is calculated as the time in minutes from the time of going to bed to the time of first nocturnal void, or time of awakening if no void occurred. The FUSP for a visit is calculated as the average of the single-night values. Aggregated values are derived similarly to the derivation of the aggregated values for number of nocturnal voids.

The total sleep for a single night is calculated as the time in minutes from the time of going to bed to the time of awakening. The total sleep for a visit is calculated as the average of the single-night values.

The diuresis rate measured in mL/min for a single night is calculated as NUV/total sleep. The diuresis rate for a visit is calculated as the average of the single-night values.

The NPi for a single night is calculated as 100%·NUV/24-hours urine volume. The NPi for a visit is calculated as the average of the single-night values.

The NI Diary outcomes will be scored as described in Holm-Larsen (2014):
The NI Diary Total Score for a single night will be calculated as the sum of the scores for the first 11 questions divided by 44 and multiplied by 100. If any of the single-item scores is missing, the NI Diary Total Score is missing for that single night. The NI Diary Total Score for a visit is calculated as the average of the single-night values. The number of nights on which the NI Diary Total Score for a visit is based will be provided.
The NI Diary Overall Impact Score for a single night will be calculated as the score for the 12th question divided by 4 and multiplied by 100. The NI Diary Overall Impact Score for a visit is calculated as the average of the single-night values. The number of nights on which the NI Diary Overall Impact Score for a visit is based will be provided.
Aggregated values for NI Diary Total Score as well as NI Diary Overall Impact Score are derived similarly to the derivation of the aggregated values for number of nocturnal voids.

The PGI-S and the PGI-I are scored as described in Viktrup (2012):
The PGI-I is scored from 1 (*Very much better*) to 7 (*Very much worse*).
The PGI-S is scored from 1 (*None*) to 4 (*Severe*).

The Hsu 5-point Likert Bother scale is scored as described in Hsu (2015):
The Hsu Bother scale is scored from 0 (*Not at all*) to 4 (*Extremely*).

The ISI will be scored as described in Bastien (2001):
For the ISI, each item is scored from 0 (*None/Very satisfied/Not interfering at all/Not at all noticeable/Not at all*) to 4 (*Very severe/Very dissatisfied/Very much interfering/Very much noticeable/Very much*) and ISI is calculated as the sum of the single-item scores. ISI is missing if any single item is missing.

10.3.3 Bayesian estimation of the sigmoidal dose-response curve

Bayesian estimation of the sigmoidal dose-response curve for aggregated, quantitative outcomes will follow the approach described for the primary analysis in Section 10.2.1 using the priors $\beta \sim N(0, 10^2)$; $\alpha_1, \alpha_2 \sim N(\bar{a}, 5^2)$ with the outcome-specific mean $\bar{a}$ presented in Table 6; $\alpha_3 \sim N^+(2, 10^2)$; $\alpha_4 \sim N^+(5, 10^2)$; $\xi_i \sim N(0, 1)$; $\delta_i \sim N(0, \tau^2)$; and $\sigma^2, \tau^2, \lambda_i^2 \sim IG(0.5, 0.01)$; after rescaling the variables using the outcome-specific scaling factors presented in Table 6. The baseline will be subtracted by the constant presented in Table 6.

Table 6 Constant to be subtracted from baseline, mean in the prior distributions of α_1 and α_2, and scaling factor used in the Bayesian analyses

Efficacy outcome	Constant to be subtracted from baseline	Mean $\bar{a}$ of the prior distributions of $\alpha 1$ and $\alpha 2$	Scaling factor
NUV	750 (mL)	−1	250 (~¼ L)
FUSP	180 (minutes)	1.5	60 (~1 hour)
NI Diary Total Score	50 (percentage points)	−2	10 (~10 percentage points)
NI Diary Overall Impact Score			

For each efficacy outcome, the posterior probability that FE 201836 performs better than placebo will be presented together with the difference to placebo with 95% credibility intervals for each dose of FE 201836.

Bayesian estimation of the sigmoidal dose-response curve for the aggregated responder variable corresponding to 50% reduction from baseline-RT in number of nocturnal voids, $z_{50\%,4,i}$, will follow the primary analysis approach with imputed outcomes derived from the imputed values of $y_{4,i}$ from the primary analysis. The dose-response model for $z_{50\%,4,i}$ will be modelled as a binary variable with mean

$$\text{Mean}(z_{50\%,4,i}) = \text{expit}\left((-0.5 \cdot x_{0,i} - \text{Mean}(y_{4,i}))/\text{SD}(y_{4,i}) \right),$$

where expit is the function expit(x)=exp(x)/(1+exp(x)), Mean($y_{4,i}$)=$\beta \cdot (x_{0,i} - 2) + \mu_d$, and SD($y_{4,i}$)=$\sigma$ as specified in (1). The model parameters are modelled with the same prior distributions as used in the primary analysis. The posterior probability that FE 201836 performs better than placebo will be presented together with the odds ratio of at least 50% reduction from baseline-RT in number of nocturnal voids during 12 weeks of treatment for each dose of FE 201836 relative to placebo with 95% credibility intervals.

10.3.4 Treatment effects on cross-sectional change from baseline for quantitative variables and PGI-I

Treatment effects on cross-sectional change from baseline for quantitative variables will be analysed using SAS PROC MIXED assuming a MMRM (a non-parametric dose-response curve is assumed). The systematic effects will comprise visit number and treatment-by-visit as categorical variables, and the baseline-RT value as a linear covariate interacting with visit number. An unstructured covariance matrix will be used if technically possible, alternatively a Toeplitz, a compound-symmetry, or if necessary, an independent covariance matrix will be used. The treatment contrasts (with associated 95% confidence intervals without multiplicity adjustment) versus placebo will be tabulated and graphically illustrated. The analyses will be applied to the observed data on the ITT-RT analysis set.

The same approach will be applied to the aggregated average change from baseline in number of nocturnal voids.

For PGI-I there is no baseline so the systematic effects will comprise only visit number and treatment-by-visit as categorical variables.

10.3.5 Treatment effects on cross-sectional binary variables

Binary endpoints such as the cross-sectional "responder status" at a visit will be analysed using SAS PROC GENMOD with logit as the link function for the responder probability. An unstructured correlation matrix will be used as working correlation matrix if technically possible, alternatively an autoregressive, a compound-symmetry, or if necessary, an independent working correlation matrix will be used. The systematic effects will include visit and treatment-by-visit interaction as factors and will be adjusted for the $\log_2$-transformed number of nocturnal voids at baseline-RT, subtracted by 1, as a covariate interacting with visit number. Odds ratios relative to placebo with associated 95% confidence intervals (without multiplicity adjustment) will be presented. The analyses will be applied to the observed data for the ITT-RT analysis set. Combinations of visit and dose of FE 201836 where all or none of the subjects respond will be omitted from the analysis since this corresponds to estimates of infinity in a logistic regression model.

For the investigation of interaction with sex, combinations of sex, visit and dose of FE 201836 where all or none of the subjects respond will be omitted from the analyses, and an autoregressive correlation matrix will be used, if technically possible, alternatively, a compound-symmetry, or if necessary, an independent working correlation matrix will be used.

In the analysis of the responder variable corresponding to one nocturnal void decrease from baseline, only subjects with at least one nocturnal void at baseline will be included. In the analysis of the responder variable corresponding to two nocturnal voids decrease from baseline, only subjects with at least two nocturnal voids at baseline will be included. In the analysis of the responder variable corresponding to three nocturnal voids decrease from baseline, only subjects with at least three nocturnal voids at baseline will be included.

10.3.6 Treatment effects on single-time-point quantitative variables

Treatment effects on single-time-point quantitative variables will be analysed using SAS PROC MIXED to perform an ANCOVA (a non-parametric dose-response curve is assumed). For the percentages of nights with response, the systematic effects will comprise treatment as a categorical variable and baseline-RT number of nocturnal voids minus 2 as a linear covariate. For change from baseline-RT in copeptin, the systematic effects will comprise treatment as a categorical variable and the baseline-RT level minus 13.1 as a linear covariate. For change from baseline-RT to lowest observed serum sodium, the systematic effects will comprise treatment as a categorical variable and the baseline-RT level minus 135 as a linear covariate. The treatment contrasts (with associated 95% confidence intervals without multiplicity adjustment) versus placebo will be tabulated and graphically illustrated. The analyses will be applied to the observed data on the ITT-RT analysis set. For change from baseline-RT to lowest observed serum sodium, the analyses will be performed both using randomized treatment for the observed data on the ITT-RT analysis set for comparability with the results for the efficacy variables and using actual treatment for the Safety-RT analysis set for comparability with the analyses of episodes of hyponatraemia based on observed serum sodium.

10.4 Therapeutic Window

The estimated sigmoidal dose-response curve with 95% credibility limits compared to placebo (left-sided y-axis) will be illustrated together with the incidence per 100 person-weeks with 95% confidence limits of observation of serum sodium <130 mmol/L and ≤125 mmol/L, respectively, at or before 12 weeks (both on the right-sided y-axis). The estimated treatment effects with 95% confidence limits compared to placebo for the groups treated with FE 201836 based on the MMRM model described in Section 10.3.4 will also be presented in the graph. The horizontal x-axis will represent the dose range of FE 201836. The therapeutic window is defined as the dose range where FE 201836 is deemed efficacious while the incidence of hyponatraemia is acceptable. Two illustrations will be made with different estimated dose-response curves for the efficacy of FE 201836: one using the results from the primary analysis and one using Sensitivity Analysis 6 based on the actual treatment received.

The sex-specific therapeutic windows will be illustrated similarly.

The therapeutic window of nocturnal volume will be presented similarly.

In addition, similar illustrations of the therapeutic window will be produced using the estimated treatment contrast compared to placebo for the change from baseline-ENR to lowest post-baseline measurement of serum sodium as the safety indicator. To make the results for the efficacy indicator

and the safety indicator comparable in each of these figures, the same treatment definition and the same analysis set will be used for both indicators in each figure.

10.5 Explorative Benchmarking Analysis

The explorative benchmarking analysis will descriptively summarize aggregated as well as visit-specific changes from baseline and percentage changes from baseline in number of nocturnal voids by all treatment groups, including desmopressin, for each visit in the randomized-treatment period. The summary statistics will include 95% confidence limits for simple means. No imputations will be done for this analysis which will be based on the observed cases only. Tables and figures will be produced as appropriate. This analysis will also be repeated by sex.

Similar tables will be produced for the quantitative, secondary endpoints: percentage of nights during treatment period with at most one nocturnal void, percentage of nights during treatment period with complete response, change from baseline in NUV, change from baseline in FUSP, change from baseline in nocturnal diuresis rate, change from baseline in NI Diary Total Score, change from baseline in NI Diary Overall Impact Score, PGI-I urinary symptoms score, change from baseline in PGI-S score, change from baseline in Bother as measured by the Hsu 5-point Likert Bother Scale, change from baseline in ISI.

Frequency tables by treatment for each visit will be produced for the categorical, secondary endpoints: response defined as 50% reduction in nocturnal voids from baseline based on aggregated average change from baseline, and response defined as 50% reduction in nocturnal voids from baseline based on visit-specific change from baseline.

11 Safety

11.1 General Considerations

11.1.1 Analysis Sets

Safety analyses with respect to the enrichment period will be based on the Safety-ENR analysis set, while analysis with respect to the randomized-treatment period will be performed using the Safety-RT analysis set. For calculation of percentages, the denominator will be the total number of subjects in the treatment group in the corresponding safety analysis set.

All analyses will be based on the actual treatment received.

11.1.2 Imputation Rules

Missing values will be treated as missing, except for the causality, intensity, seriousness and outcome of adverse events and for the definition of treatment emergent. For adverse events (AEs), a worst-case approach will be used: if causality is missing, the adverse event will be regarded as related to the IMP; if the intensity of an adverse event is missing, the adverse event will be regarded as severe; if seriousness is missing, the adverse event will be regarded as serious; if onset date is missing, it will be assumed to be at the date of Visit 2; if outcome is missing and no date of outcome is present, the outcome is regarded as 'ongoing' at the End-of-Trial Visit.

11.1.3 Visit and Visit Windows

See Section 3.1.5 for visit windows.

11.2 Safety Endpoints of Special Interest

11.2.1 Hyponatraemia Episodes Based on Observed Serum Sodium Levels

11.2.1.1 General Considerations

All tables and figures will be done for the overall trial population and by gender.

Observed serum sodium levels will both be summarized as a continuous variable and be used to define worst-case hyponatraemia episodes using the definitions specified in Table 7.

Table 7 Definition of Hyponatraemia Episodes by Serum Sodium Range: Number of Consecutive Measurements of Serum Sodium below Upper Threshold Needed to Start Episode

Degree	Serum Sodium Range (mmol/L)	Enrichment Period	Randomized-Treatment Period	
		Hyponatraemia Episode	Hyponatraemia Episode	'Confirmed' Hyponatraemia Episode
None	≥135	Not applicable (normal serum sodium level)		
Mild	≥130 to <135	1st observation	1st observation	3rd consecutive observation
Moderate	>125 to <130	1st observation	1st observation	2nd consecutive observation
Severe	≤125	1st observation	1st observation	1st observation

11.2.1.2 Observed Serum Sodium Levels

Observed serum sodium levels and the change from baseline-ENR will be summarized by visit for the Safety-ENR analysis set. Observed serum sodium levels and the change from baseline-RT will be summarized by visit and treatment group for the Safety-RT analysis set.

All subjects in the Safety-ENR analysis set with a serum sodium value <130 mmol/L at any time point will be listed by actual treatment group for the randomized period (non-randomized will be listed as a separate treatment group), including all serum sodium assessments by time point and including demographic data.

Graphs will present the serum sodium levels against analysis day (ADY) by treatment group and the subject profiles with a serum sodium level <135 mmol/L at any time point during the trial will be highlighted. In addition, lowest observed serum sodium against age will be presented in two figures, one figure using plot symbols showing sex and one figure using plot symbols showing actual treatment group.

11.2.1.3 Hyponatraemia Episodes During the Randomized-Treatment Period based on Observed Serum Sodium Levels

When presenting occurrence (measured as number of subjects) by visit of worst-case hyponatraemia episodes as defined in Table 7, all visits within the same visit window will be combined to define the worst-case hyponatraemia episode observed for the subject within the visit window. The occurrence of worst-case hyponatraemia episodes will be presented by treatment group and visit for the randomized-treatment period for the Safety-RT analysis set.

The number of the post-baseline-ENR 'worst-case' hyponatraemia episodes for each subject will be summarized overall and by visit as a categorical variable; here 'worst-case' for each episode is defined as the worst-case serum sodium level observed during the hyponatraemia episode.

In addition, the incidence of "Hyponatraemia Episode" (Table 7) per 1000 person-years at risk counting all events for a subject will be presented overall and by visit using the upper end of the visit window specified for Category I endpoints in Table 4 as cut-points for the categorization of the at-risk time.

For each of the thresholds (<135, <130, and ≤125 mmol/L), graphs presenting treatment-stratified Kaplan-Meier curves of time from randomization (Visit 4) to first serum sodium level below the threshold will be produced for the Safety-RT analysis set. The time to event is calculated as the date of the first serum sodium level below the threshold minus the date of randomization.

In addition, for the thresholds <135 (mild to severe) and <130 mmol/L (moderate to severe), treatment-stratified Kaplan-Meier plots corresponding to first 'confirmed' hyponatraemia episode will be produced. Here the date of confirmation is used as the date of the event. For mild to severe, the date of confirmation is the first of the dates of '3rd consecutive observation <135 mmol/L', '2nd consecutive observation <130 mmol/L', and '1st observation ≤125 mmol/L'. For moderate to

severe, the date of confirmation is the first of the dates of '2nd consecutive observation <130 mmol/L' and '1st observation ≤125 mmol/L'.

In Kaplan-Meier plots, all subjects are censored at the earliest of End-of-Trial Visit and 85 days after randomization.

Since the event times are interval censored (the true time of crossing the threshold was after the time of the preceding serum sodium measurement and before or at the time of the current serum sodium measurement) and essentially grouped by the planned visits, the Kaplan-Meier plots are not quite appropriate. Therefore, event-free-time curves will also be produced based on life table methods using interval cut points derived as 1 + the largest observed event time within each of the visit windows specified for Category I endpoints in Table 4.

Duration of the hyponatraemia episode will be derived as 1 + end date of episode – date of start of episode. For each threshold separately, the duration will be listed for serum sodium episodes defined as one or more observed serum sodium levels below the threshold. The listings will contain the duration, demographics, the day of stop of treatment relative to the start of the hyponatraemia episode, and the serum sodium values observed during the episode including the serum sodium value above 135 mmol/L defining the end of the hyponatraemia episode. For episodes starting before randomization, duration will be listed by sex for all subjects in the Safety-ENR analysis set. For episodes starting after randomization, duration will be listed by sex and treatment group for all subjects in the Safety-RT analysis set experiencing an episode of hyponatraemia based on observed serum sodium levels.

Nadir of observed serum sodium during hyponatramia episodes will be illustrated by a graph including all hyponatraemia episodes and showing observed serum sodium levels against time since start of the hyponatraemia episode (time zero for an episode is defined as first time point with observed serum sodium below 135 mmol/L in the current episode).

11.2.2 24-hour Urine Volume

Observed 24-hour urine volume and the change from baseline-RT will be summarized by visit and treatment group for the Safety-RT analysis set. In addition, change from baseline-RT will be analysed for the using the MMRM approach described in Section 10.3.4.

11.3 Routine Safety Assessments

11.3.1 Adverse Events

11.3.1.1 General Considerations

Adverse events (AEs) are classified according to the latest version of the MedDRA dictionary. The version of MedDRA will be documented.

Written narratives will be issued for all serious AEs (SAE, including deaths) and AEs leading to discontinuation.

Regarding definition of treatment emergent, there are three levels for treatment emergent because of the single-blind enrichment period: pre-treatment, enrichment-treatment emergent, and randomized-treatment emergent. The rules for classification are the following:

- Any AE starting before date of Visit 2 is considered pre-treatment.
- Any AE starting on date of Visit 2 as well as any AE starting after Visit 2, but before date-time for randomization at Visit 4, are considered enrichment-treatment emergent.
- If the subject is not randomized, any AE starting on date of Visit 2 as well as any AE starting after Visit 2 are considered enrichment-treatment emergent.
- Any AE starting after the date-time of randomization at Visit 4 is randomized-treatment emergent.
- AEs with missing start time but starting on date of Visit 4 are considered randomized-treatment emergent.
- AEs with missing start date are considered enrichment emergent.

The total number of subjects reporting an AE, and the percentage of subjects (%) reporting an AE (denominator: number of subjects screened for pre-treatment; number of subjects in the Safety-ENR analysis set for enrichment-treatment emergent; and number of subjects in the Safety-RT analysis set for randomized-treatment emergent) will be presented.

11.3.1.2 Tabulations of Treatment-Emergent Adverse Events

Treatment-emergent adverse events will be summarized by system organ class (SOC) and preferred term (PT) and by treatment group. The tables will display the total number of subjects reporting an AE, the percentage of subjects (%) with an AE, and the number of events reported. AEs will be presented by SOC sorted alphabetically and PT sorted in decreasing frequency of occurrence.

Summary tables will be prepared for the enrichment period and randomized-treatment period as outlined in Table 8.

Table 8 Tabulations of Treatment-Emergent Adverse Events

	Enrichment Period Enrichment Period-Emergent Adverse Events (Safety-ENR Analysis Set)	**Randomized Period** Randomized-Treatment-Emergent Adverse Events (Safety-RT Analysis Data Set)	Gender-Specific Tabulations
Overview of Adverse Events by treatment group	X	X	
Adverse Events by causality (related/unrelated)	X	X	X
Adverse Events leading to death	X	X	
Adverse Events by intensity		X	X
SAEs	X	X	
Adverse Events leading to discontinuation	X	X	
Adverse Events with an incidence of at least 10% in any treatment group		X	X

	Enrichment Period Enrichment Period-Emergent Adverse Events (Safety-ENR Analysis Set)	**Randomized Period** Randomized-Treatment-Emergent Adverse Events (Safety-RT Analysis Data Set)	Gender-Specific Tabulations
Adverse drug reactions with an incidence of at least 10% in any treatment group		X	X

Supporting data listings will be provided for:

- All adverse events sorted by site and subject number
- All adverse events sorted by MedDRA Preferred Term
- Serious adverse events
- Adverse events leading to death
- Adverse events leading to discontinuation

11.3.2 Safety Laboratory Variables

11.3.2.1 General Considerations

Laboratory variables will be grouped under "Haematology", "Clinical Chemistry", "Urinalysis", or "Coagulation Factors".

The reference ranges for laboratory values are presented in Appendix 1.

11.3.2.2 Visits and Visit Windows

- Baseline for all laboratory analyses will be the values obtained at the last assessment before the subject received the first dose of the IMP.
- Treatment-emergent for laboratory data will be defined as specified for AEs in Section 11.3.1.1.
- End of trial will include the last post-baseline-ENR observation during the trial.
- Data will be analysed by planned visit using the visit-window structure as described in Section 3.1.5.

11.3.2.3 Summary Statistics

For each laboratory variable, summary statistics (see Section 1.3.1) by treatment group will be provided for test results and their change from baseline-ENR (quantitative variables only) at each visit up to and including the End-of-Trial Visit for the Safety-ENR analysis set; and for test results and their change from baseline-RT (quantitative variables only) at each visit up to and including the End-of-Trial Visit for the Safety-RT Analysis Set. The number and percent of subjects with clinically significant laboratory abnormalities according to the investigator's assessment of the categories 'normal', 'abnormal, not clinically significant', and 'abnormal, clinically significant' will be presented. For Visits 2, 3, and 4, the Safety-ENR analysis set will be used; for post-randomization visits, the Safety-RT analysis set will be used.

11.3.2.4 Laboratory Variable Changes Relative to Normal Range

Shift tables will be presented on an as-needed basis only.

11.3.3 Vital Signs

11.3.3.1 Visits and Visit Windows

- Baseline for all vital signs analyses will be the values obtained at the last assessment before the subject received the first dose of the IMP during the enrichment period.
- Treatment-emergent vital signs data will be defined as specified for AEs in Section 11.3.1.1.
- End of trial will include the last post-baseline observation during the trial.
- Data will be analysed by planned visit using the visit-window structure as described in Section 3.1.5.

11.3.3.2 Summary Statistics

For each vital sign variable, summary statistics (see Section 1.3.1) will be presented for observed values as well as the investigator's assessment of the categories 'normal', 'abnormal, not clinically significant' and 'abnormal, clinically significant'. For Visits 2, 3, and 4, the Safety-ENR analysis set will be used; for post-randomization visits, the Safety-RT analysis set will be used.

11.3.4 ECGs

- ECG will be categorized as 'normal', 'abnormal, not clinically significant', and 'abnormal, clinically significant'.
- Baseline for ECG is the assessment at the Screening Visit.
- Randomized-treatment-emergent ECG is the ECG at the End-of-Trial Visit.

The ECG variable will be summarized overall for the Safety-ENR analysis set and by treatment for the Safety-RT analysis set. Shift tables will be presented on an as-needed basis only.

11.3.5 Other Safety Variables

11.3.5.1 Weight Changes

- Weight Change will be derived as weight at the current visit minus weight at the previous visit using the visit-window structure described in Section 3.1.5.
- Data will be analysed by planned visit using the visit-window structure as described in Section 3.1.5.
- Enrichment-treatment-emergent weight changes comprise weight changes from Visit 2 to Visit 3 and weight changes from Visit 3 to Visit 4. Post-randomization weight changes are considered randomized-treatment-emergent weight changes
- The end of trial value for a subject will include the last post-baseline-ENR observation for the subject during the trial.

Summary tables will be presented by visit displaying the number and percentage of subjects with a weight change of more than 2 kg since the previous visit.

11.3.5.2 Physical examination

- Physical examination findings will be categorized according to 'normal', 'abnormal, not clinically significant', and 'abnormal, clinically significant'.
- Baseline-ENR for the physical examination is the assessment at the Screening Visit, baseline-RT is the assessment at Visit 4, and End of Trial will include the last post-baseline-RT assessment during the trial.

Summary tables will be prepared by visit for the Safety-ENR analysis set, and by visit and treatment for the Safety-RT analysis set. Shift tables will be presented on an as-needed basis only.

12 Interim analyses

The safety data will be reviewed on a regular basis by a Data Monitoring Committee (DMC) as described in a separate DMC charter.

After 125 subjects have been randomized and every 8 weeks thereafter, an interim analysis is performed by [redacted]. Each of these interim analyses comprises a futility analysis and, until end of recruitment, an update of allocation probabilities for the FE 201836 doses based on a Bayesian analysis of the dose-response relation. The updating of the allocation probabilities is focused on improving the design with regard to the identification of the MED and the ED_{85}.

In contrast to the primary analysis, the interim analyses will not be adjusted for baseline number of nocturnal voids.

In the first interim, the futility stopping rule is to stop for futility if $Pr(\mu_{500} - \mu_{Plb} < -0.3) < 10\%$. In the later interim analyses, the futility rule is to stop for futility if $Pr(\mu_{500} - \mu_{Plb} < -0.3) < 5\%$. At all these trial updates, Bayesian estimates will be used to update the allocation probabilities for the doses of FE 201836. The allocation probability for each dose *d* of FE 201836 is proportional to the posterior variance of μ_d multiplied by the average of Pr(*d* is the MED) and Pr(*d* is the ED_{85}), and divided by $n_d + 1$.

The power and the probability of making a false positive conclusion using the adaptive design was investigated based on simulations as described in the trial protocol.

13 Risk-benefit assessment

Risk-benefit will be evaluated based on the graphical illustrations of the therapeutic window (Section 10.4).

14 Deviations From the Trial Protocol

A Full Analysis Set was defined in the protocol. However, none of the analyses specified in the trial protocol were based on the Full Analysis Set, so no Full Analysis Set has been defined in this SAP.

$Pr(\mu_d<\mu_{Plb})$ will presented only for the 500 µg FE 201836 dose (Proof of Concept) since all doses will have the same probability because of the monotone dose-response curve parameterization; and $Pr(\mu_d=\mu_{MAX})$ will not be displayed since it always equals 1 for the 500 µg dose, and probabilistically almost surely (that is, with probability 1) equals 0 for all other doses, again because of the monotone dose-response curve parameterization.

For some subjects, the baseline voiding diary is missing due to technical problems in the beginning of the study. The change from baseline in voiding data are considered missing completely at random for these subjects, and the subjects are not included in any analyses involving baseline measurements of voiding data.

Two sensitivity analyses have been added:

- Estimates of all quantities of interest based on the Bayesian analysis of the longitudinal model will be presented.
- A Bayesian analysis of the primary outcome with imputation using the level corresponding to placebo for drop-out.

The first sensitivity analysis presents the result for the model used for imputation. The second analysis is the Bayesian equivalent to the frequentist approach specified in the trial protocol based on a mixed model for repeated measurement (repeated-measurements ANCOVA) using MNAR placebo-based pattern mixture model.

Linear trend estimates have been added to the above mentioned protocol-specified frequentist approach as supplements to the description of the association with dose in case there are difficulties with the convergence for the primary analysis or very large uncertainties on the estimated α parameters.

A secondary analysis of aggregated average change from baseline in nocturnal urine volume has been added. This outcome variable was added to comply with the Bayesian analysis mentioned in Table 9-4 in the trial protocol.

The ITT-RT data set will be used for all secondary analyses as stated in the protocol; but only observed data will be used. This is considered a clarification of the protocol.

For all baseline-adjusted analyses of more than one visit at a time, the interaction between visit and baseline has been added as a systematic effect. This is done because it is unlikely that the baseline value has the exact same influence on measurements 1 week after baseline and 12 weeks after baseline. Furthermore, this leads to a frequentist approach more similar to the Bayesian model in which the association with baseline depends on the visit. For all analyses estimating the interaction between treatment and sex, interactions have been added between sex and each of the systematic

effects. This is done to prevent potentially contradictory results from separate analyses of the two sexes and the evaluation of the interaction between treatment and sex (overall or visit-specific, as appropriate) caused by non-comparable statistical models.

The protocol states in Section 9.7.3 Secondary and Exploratory Analyses:
Binary endpoints such as "responder status" at a certain visit will be analysed using SAS PROC GENMOD with the logit as link function for the responder probability. An unstructured covariance matrix will be assumed for the repeated responder status which will be adjusted for baseline value as covariate and treatment, visit and treatment-by-visit interaction as factors. (Associated with e.g., responder analysis in terms of reduction in number of voids, cumulative distritubtion function plots of percentage change from baseline by treatment may be produced.)
This has been clarified to the following:
Binary endpoints such as the cross-sectional "responder status" at a visit will be analysed using SAS PROC GENMOD with logit as the link function for the responder probability. An unstructured correlation matrix will be used as working correlation matrix. The systematic effects will include treatment, visit, and treatment-by-visit interaction as factors, and will be adjusted for the $\log_2$-transformed number of nocturnal voids at baseline-RT, subtracted by 1, as a covariate interacting with treatment. Odds ratios relative to placebo with associated 95% confidence intervals (without multiplicity adjustment) will be presented. The analyses will be applied to the observed data for the ITT-RT analysis set.

In addition to the illustrations of the therapeutic window specified in the protocol, illustrations of the therapeutic window will be produced using the estimated treatment contrast compared to placebo for the change from baseline-RT to lowest post-baseline measurement of serum sodium as the safety indicator. The inclusion of desmopressin in the plots illustrating the therapeutic window ("Corresponding efficacy and safety of desmopressin will be illustrated in a side panel for comparison.") has been deleted to comply with "None of the primary, secondary or explorative analyses will include desmopressin." (Section 10.1).

In the analyses of time to 'confirmed' hyponatraemia episodes, the time to event is defined as the time of confirmation instead of the time to first measurement below the threshold to avoid conditioning on the future. The protocol states "The associated relative hazards of treatment and gender will be analysed using Cox proportional hazards model adjusting for baseline serum sodium and age (<65, ≥65 years). Possible heterogeneity between the genders in the relative treatment hazards will be explored by the inclusion of a treatment-by-gender interaction term". These analyses will not be performed since the trial is too small to perform Cox regression analyses of the occurrence of hyponatraemia events with 9 treatment groups and 2 covariates.

The adverse events tables presenting incidences of least 5% have been changed to presenting at incidences of at least 10% because it is expected that for several of the FE 201836 treatment groups, 20 subjects or less are allocated to the group.

According to [redacted], the interim analyses will deviate from the Appendix 2 in the protocol regarding the derivation of the allocation probabilities by the factor $1/(n_d + 1)$ before the normalization to sum to 1.

Remaining changes are all considered clarifications of presentations specified in the protocol.

15 Tables, Listings and Figures

Non-standard table and listing shells and figure specifications will be provided in a separate document on an as-needed basis.

Standard tables and listing are according to the end-of-text and in-text specifications specified at O:\GB_Prog\05_Data Standard\Standardisation of output.

16 References

Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med. 2001 Jul; 2 (4):297-307.

Holm-Larsen T, Andersson F, van der Meulen E, Yankov V, Rosen RC, Nørgaard JP. The nocturia impact diary: a self-reported impact measure to complement the voiding diary. Value Health. 2014 Sept; 17 (6): 696-706.

Hsu A, Nakagawa S, Walter LC, Van Den Eeden SK, Brown JS, Thom DH, et al. The Burden of Nocturia Among Middle-Aged and Older Women. Obstet Gynecol. 2015; 125 (1): 35-43.

Ratitch B, O'Kelly M. Implementation of Pattern Mixture Models Using Standard SAS/STAT Procedures, PharmaSUG2011 – Paper SP04

Viktrup L, Hayes RP, Wang P, Shen W. Construct validation of patient global impression of severity (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract symptoms secondary to benign prostatic hyperplasia. BMC Urology. 2012; 12 (1): 1-8.

Appendices

Appendix 1 Markedly Abnormal Laboratory Safety Values, Vital Signs and ECGs

Table A: Criteria for Markedly Abnormal Laboratory Test Results (Haematology)

Variable	Units	Criteria for Markedly Abnormal	
		Low	High
Haemoglobin	g/L	≤ 80	≥ 200
Haematocrit	Ratio	≤ 0.30	≥ 0.54
Total WBC	10^9/L	≤ 2.0	≥ 35.0
Total RBC	10^{12}/L	≤ 3.0	≥ 6.2
MCV	fL	≤ 75	≥ 105
MCH	pg	≤ 25	≥ 35
MCHC	mmol/L	≤ 18.62	≥ 22.96
Basophils	%	≤ 0	≥ 1.5
Absolute Basophils	10^9/L	≤ 0	≥ 2.0
Eosinophils	10^9/L	< 3.0	Not applicable
Leucocytes	10^9/L	< 2.5	> 15.0
Lymphocytes	10^9/L	< 0.50	> 20.00
Monocytes	%	< 3.4	> 9.0
Absolute Neutrophil Count	10^9/L	< 1	Not applicable
Platelets	10^9/L	< 50	> 999

MCH: Mean corpuscular haemoglobin
MCHC: Mean corpuscular haemoglobin concentration
MCV: Mean corpuscular volume
RBC: Red blood cells
WBC: White blood cells

Table B: Criteria for Markedly Abnormal Laboratory Test Results (Clinical Chemistry)

Variable	Units	Criteria for Markedly Abnormal	
		Low	High
AST (SGOT)	IU/L	Not applicable	> 5 x ULN AST(female): >155 AST(Male): >185
ALT (SGPT)	IU/L	Not applicable	> 5 x ULN ALT (female): >165 ALT(Male): >205
Alkaline phosphatase	IU/L	Not applicable	5 x ULN ALP (female): > 520 ALP(Male): >645
GGT	IU/L	Not applicable	> 3 x ULN
LDH	IU/L	Not applicable	> 3 x ULN
Globulin	g/dL		
Total bilirubin	mg/dL	Not applicable	≥ 1.5 x ULN
Direct bilirubin	mg/dL	Not applicable	≥ 8.55
Urea nitrogen	mmol/L	Not applicable	≥ 10.7
Creatinine	umol/L	Not applicable	Female: >168 Male: > 206
Glomerular Filtration Rate	mL/min	≤ 60	Not applicable
Total protein	g/L	≤ 45	≥ 90
Albumin	g/L	≤ 20	≥ 65
Potassium	mmol/L	≤ 2.5	≥ 6.5
Sodium	mmol/L	≤ 125	> 155
Chloride	mmol/L	≤ 90	≥ 115
Phosphorus	mmol/L	≤ 0.5	≥ 1.9
Calcium	mmol/L	≤ 1.75	≥ 3.13
Bicarbonate	mmol/L	≤ 15	≥ 34
Glucose	mmol/L	≤ 2.8	> 13.9
Magnesium	mmol/L	≤ 0.6	≥ 1.1
Total cholesterol	mmol/L	Not applicable	≥ 10.36
Triglycerides	mmol/L	Not applicable	> 5.65
CPK	U/L		> 850

ALT: Alanine transaminase
ALP: Alkaline phosphatase

AST: Aspartate transaminase
CPK: Creatine phosphokinase
GGT: Gamma glutamyl transferase
IU:
LDH: Lactate dehydrogenase
SGOT: Serum glutamic oxaloacetic transaminase
SGPT: Serum glutamic pyruvic transaminase
ULN: Upper Limit Normal

Table C. Criteria for Markedly Abnormal Vital Sign Values* and Body Weight

Variable	Values	Change from Baseline
Systolic blood pressure	≥ 180 mmHg ≤ 90 mmHg	Increase of ≥ 20 mmHg Decrease of ≥ 20 mmHg
Diastolic blood pressure	≥ 105 mmHg ≤ 50 mmHg	Increase of ≥ 15 mmHg Decrease of ≥ 15 mmHg
Pulse rate	≥ 120 bpm ≤ 50 bpm	Increase of ≥ 15 bpm Decrease of ≥ 15 bpm
Body Weight	None	Increase of ≥ 7% or 2 kg between trial visits Decrease of ≥ 7%

Bmp: beats per minute

* To be considered as markedly abnormal, a value must be above or below the specified values and fulfill one of the criteria for change from baseline (increase or decrease).