Document Type:	Statistical Analysis Plan
Official Title:	A double-blind, randomized, placebo-controlled multicenter study to investigate efficacy and safety of elinzanetant for the treatment of vasomotor symptoms over 26 weeks in postmenopausal women
NCT Number:	NCT05099159
Document Date:	23 OCT 2023

Title Page

Protocol Title: A double-blind, randomized, placebo-controlled multicenter study to investigate efficacy and safety of elinzanetant for the treatment of vasomotor symptoms over 26 weeks in postmenopausal women.

Protocol Number: 21652

Compound Number: BAY 3427080 / elinzanetant

Short Title: Overall **A**ssessment of efficacy and **S**afety of elinzanetant **I**n patients with vasomotor **S**ymptoms (OASIS-2)

Acronym: OASIS 2

Sponsor Name: Bayer Consumer Care AG

Legal Registered Address: Peter-Merian-Strasse 84, 4052 Basel, Switzerland

Regulatory Agency Identifier Number(s):

Registry	**ID**
EudraCT	2020-004855-34

Date: 23 OCT 2023
Version: 3.0

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.
The approval of the Statistical Analysis Plan is documented in a separate signature document.

Table of Contents

Table of Tables

Table of Figures

Version History

This Statistical Analysis Plan (SAP) for study 21652 is based on the protocol Version 4.0 dated 22 JUN 2022.

SAP Version	Date	Change	Rationale
1.0	06 JUN 2023	Not Applicable	Original version
2.0	17 OCT 2023	This amendment of the SAP includes correction of typos, clarifications of wording and analyses as well as additions of data handling rules. Major changes have been incorporated in following sections:	Needed changes were identified during review of this document and of blinded data.
		1) Number of decimal places for standard deviation (see Section 5.1 General Considerations) was changed from 2 to 1 place more than in original data. 2) Descriptive display of change from baseline was changed from boxplots to line plots for means together with 95% CIs.	Changes were made based on blinded review of tables and figures and to be in line with Bayer's programming standard.
		3) Data handling for participants who were randomized but never started treatment was clarified (Section 5.1.2). 4) Data handling of ePRO assessments for participants after premature treatment discontinuation was clarified (Section 5.1.2).	Clarification of data handling based on blind data review
		5) Modifications to description of data handling with respect to the supplementary analyses and missing data imputation strategy in Section 5.3.4 and in 6.3.	Modifications introduced to improve the clarity of the data handling rules without changing the underlying assumptions and analysis strategy

		6) The liver monitoring for the laboratory parameter INR was updated to uses the absolute value instead of the relation to the upper limit of normal as described in Section 5.6.3.4.	Update according to the FDA guidance (2009)
		7) In Sections 6.5.2 and 6.5.3, the preliminary list of alternative VMS treatment and the preliminary list of prohibited concomitant medication were updated. Two tables were added to allow a specification in addition by drug names.	Update based on blind data review
		8) In Section 6.7, the AESI term "Photosensitivity" has been changed to "Phototoxicity" and the applicable coding conditions have been updated to "BMQ Photosensitivity reactions". 9) The search for AESI "Any condition triggering close liver observation" was clarified. 10) For AESI "Post-menopausal uterine bleeding" typos were corrected and a new entry titled "PT Unexpected vaginal bleeding on hormonal IUD" has been added.	Update of applicable coding conditions
3.0	23 OCT 2023	This amendment of the SAP includes a correction of the p-value that will be reported based on the non-parametric rank ANCOVA in the sensitivity analyses (see Section 5.3.3.1). We will report a two-sided (instead of a one-sided) p-value, i.e., the two-sided alternative hypothesis is tested).	Only a two-sided test is available in SAS.

1. Introduction

Vasomotor symptoms (VMS), commonly referred to as hot flashes (HF) (also called "flushes"), are one of the most common, bothersome and distressing symptoms felt by women during the menopause transition, and the leading cause for seeking medical attention during this particular phase of a woman's life (Pachman et al. 2010). Effective treatment options are mostly limited to hormone therapy, which despite being effective is associated with some safety concerns such as increased risk of hormone-dependent cancers and cardiovascular adverse effects, including thrombotic risk (BIJUVA® Prescribing Information 2018, PREMARIN® Prescribing information 2017).

Since hormone treatment is not an option for many women suffering from moderate and severe vasomotor symptoms, either because of contraindications or personal preferences, there is a strong medical need for an effective non-hormonal treatment option with improved benefit / risk balance.

Elinzanetant is a dual neurokinin-1,3 receptor antagonist. Emerging data indicate that hot flashes may be treated by targeting the neuroendocrine factors that trigger the vasomotor symptoms (Rance et al. 2013).

Two previous studies, RELENT-1 (NCT02865538 2020) and SWITCH-1 (814-PM-02) (NCT03596762 2020), have shown elinzanetant to be significantly better than placebo in reducing the frequency and severity of hot flashes. Furthermore, the SWITCH-1 study showed that the reduction in frequency and severity of hot flashes was associated with marked improvements on participant reported outcomes of sleep, mood and quality of life.

It is anticipated that elinzanetant will be a relevant improvement compared to established available non-hormonal therapies. The efficacy is expected to be comparable to that seen with hormone therapy, but with a quicker onset of action and without the warnings and precautions and most contraindications of hormonal therapy. It is also anticipated to result in greater efficacy and a better tolerability profile than anti-depressants used for treating vasomotor symptoms.

Study 21652 is one of several Phase 3 studies, aimed at characterizing the efficacy and safety of elinzanetant for the treatment of vasomotor symptoms related to menopause.

The SAP describes the final analysis of the study. No statistical interim analysis will be performed. Table, figure and listing specifications are contained in a separate document.

1.1 Objectives, Endpoints and Estimands

Objectives and endpoints are listed in Table 1–1.

Based on available FDA (FDA 2003) and EMA (EMA 2005) guidance on the treatment of VMS associated with menopause the requirements regarding the primary endpoints for efficacy trials are different. Therefore, following the ICH E17 guidance on multiregional trials (EMA 2017), a differing set of primary and key secondary endpoints have been defined for the two regions. Details can be found in the table below.

Table 1–1: Objectives and endpoints

Objectives	Endpoints
Primary	
• To evaluate the efficacy of elinzanetant for the treatment of VMS associated with the menopause	**<u>For regulatory submission in the US:</u>** **Primary endpoints** • Mean change in frequency of moderate to severe HF from baseline to Week 4 (assessed by HFDD) • Mean change in frequency of moderate to severe HF from baseline to Week 12 (assessed by HFDD) • Mean change in severity of moderate to severe HF from baseline to Week 4 (assessed by HFDD) • Mean change in severity of moderate to severe HF from baseline to Week 12 (assessed by HFDD) **<u>For all other regions except regulatory submission in the US:</u>** **Primary endpoints** • Mean change in frequency of moderate to severe HF from baseline to Week 4 (assessed by HFDD) • Mean change in frequency of moderate to severe HF from baseline to Week 12 (assessed by HFDD) **Key secondary endpoints** • Mean change in severity of moderate to severe HF from baseline to Week 4 (assessed by HFDD) • Mean change in severity of moderate to severe HF from baseline to Week 12 (assessed by HFDD) **<u>For all regions:</u>** **Exploratory endpoints:** • Proportion of participants with at least 50% reduction in frequency of HF at week 4 • Proportion of participants with at least 50% reduction in frequency of HF at week 12
Secondary	
• To evaluate the onset of efficacy of elinzanetant for the treatment of VMS associated with the menopause	**Key secondary endpoints** • Mean change in frequency of moderate to severe HF from baseline to Week 1 (assessed by HFDD) **Secondary endpoints:** • Mean change in frequency of moderate to severe HF from baseline over time (assessed by HFDD) **Exploratory endpoints:** • Time to treatment response* • Mean change in frequency of mild, moderate, and severe HF from baseline over time (assessed by HFDD)**

Objectives	Endpoints
• To evaluate the efficacy of elinzanetant in women treated for relief of VMS associated with the menopause on: ○ sleep quality ○ menopause related quality of life ○ depressive symptoms	**Key secondary endpoints:** • Mean change in PROMIS SD SF 8b total score from baseline to Week 12 • Mean change in MENQOL total score from baseline to Week 12 **Secondary endpoints:** • Mean change in BDI-II total score from baseline to Week 12 • Mean change in BDI-II total score from baseline to Week 26 **Exploratory endpoints:** • Absolute values and changes in the ISI total score over time • Absolute values of the PGI-C individual item scores over time • Absolute values and change in PGI-S individual item scores over time • Absolute values and change in EQ-5D-5L single dimensions and health state VAS score over time • Absolute values and changes in the BDI-II total score over time • Mean change in MENQOL domain and single item scores from baseline over time** • Mean change in frequency of nighttime awakening from baseline over time (assessed by HFDD)** • Mean change in proportion of days with participants rating of "quite a bit" or "very much" sleep disturbances experienced due to HF from baseline over time (assessed by HFDD)**
• To evaluate the safety of elinzanetant for the treatment of VMS associated with the menopause	• Number of participants with TEAEs • Mean change in sleepiness score assessed by Sleepiness Scale at Week 1, Week 4, and Week 12 compared to baseline
Other pre-specified	
• To evaluate variability in exposure in relation to the efficacy and safety for elinzanetant	• Systemic exposure of elinzanetant in plasma via sparse PK sampling
• To further investigate elinzanetant (e.g., mode-of-action-related effects, safety) and to further investigate pathomechanisms deemed relevant to VMS and associated health problems	• Various biomarkers (e.g., diagnostic, safety, pharmacodynamic, monitoring, or potentially predictive biomarkers)

BDI-II=Beck Depression Inventory, EQ-5D-5L=European Quality of Life 5-dimension 5-level questionnaire, HF = Hot Flash, HFDD=Hot Flash Daily Diary, ISI = Insomnia severity index, MENQOL=Menopause Specific Quality of Life Scale, PGI-C=Patient Global Impression of Change, PGI-S=Patient Global Impression of Severity, PK = Pharmacokinetics, PROMIS SD SF 8b=Patient-reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b, TEAE = treatment emergent adverse event, US = United States, VAS = Visual analog scale, VMS = Vasomotor symptoms

*Please see Section 5.5.1 for the definition of treatment response.

**Additional exploratory endpoints.

Estimands

The attributes of the main estimand for the primary endpoints are as follows:

- Population: Post menopause women aged 40-65 with VMS as described by the inclusion/exclusion criteria detailed in the protocol.
- Variable: Efficacy will be assessed based on 2 or 4 primary endpoints depending on the region (regulatory submission in the US only/all regions) as listed below:
    - Change in frequency of moderate to severe HF from baseline to Week 4 (all regions).

- Change in frequency of moderate to severe HF from baseline to Week 12 (all regions).
- Change in severity of moderate to severe HF from baseline to Week 4 (regulatory submission in the US only).
- Change in severity of moderate to severe HF from baseline to Week 12 (regulatory submission in the US only).

- Treatment: 120 mg elinzanetant, Placebo
- Intercurrent Events (ICEs): see Table 1–2 (for further details regarding the identification of ICEs see Section 6.5)

Table 1–2: Main Estimand: ICEs and Strategies to address them

ICEs**	Reason for ICE	Strategy	Data handling method
Temporary Treatment interruption*	AEs (treatment related/unrelated)	**Treatment policy**	Utilise the collected data after ICE.
	COVID-19 and administrative reasons	**Treatment policy**	Utilise the collected data after ICE.
Permanent discontinuation of randomized treatment	AEs (treatment related/unrelated) or Lack of efficacy		
	For participants who remained untreated/on background therapy.	**Treatment policy**	Utilise the collected data after ICE.
	For participants who initiate alternative VMS treatment	**Treatment policy**	Utilise the collected data after ICE.
	Other treatment-unrelated reasons, including COVID-19	**Treatment policy**	Utilise the collected data after ICE.
Intake of prohibited concomitant medication having impact on efficacy	All reasons	**Treatment policy**	Utilise the collected data after ICE.

ICE=Intercurrent event, AE=Adverse event
*Definition of temporary treatment interruption:
Week 1 = Treatment taken on <5/7 days during week 1 (day 2 - 8).
Week 4 = Treatment taken <80% during weeks 1-4 (day 1 - 28) OR treatment taken on <5/7 days during either week 3 or 4.
Week 8 = Treatment taken <80% during weeks 1-8 (day 1 – 56) OR treatment taken on <5/7 days during either week 7 or 8.
Week 12 = Treatment taken <80% during weeks 1-12 (day 1 – 84) OR treatment taken on <5/7 days

during either week 11 or 12.
**ICEs will be reviewed at the Blinded Review Meeting prior to the study unblinding

- Population level summary:
    - Mean change in frequency of moderate to severe HF from baseline to Week 4 (all regions).
    - Mean change in frequency of moderate to severe HF from baseline to Week 12 (all regions).
    - Mean change in severity of moderate to severe HF from baseline to Week 4 (regulatory submission in the US only).
    - Mean change in severity of moderate to severe HF from baseline to Week 12 (regulatory submission in the US only).

  Treatment comparison will be based on differences in treatment group means for each endpoint.

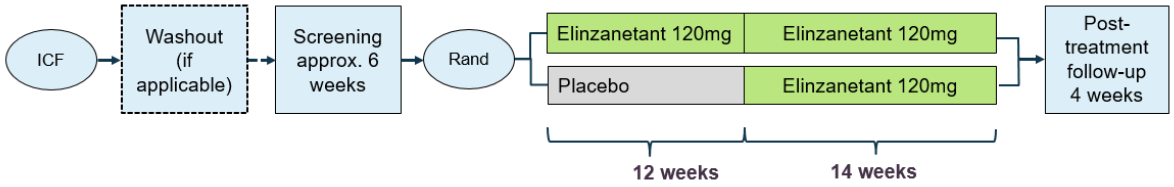
The key secondary endpoints will be handled using similar attributes except for the variables and population summary that are listed below:

- Variable:
    - Change in severity of moderate to severe HF from baseline to Week 4 (all regions except regulatory submission in the US)
    - Change in severity of moderate to severe HF from baseline to Week 12 (all regions except regulatory submission in the US)
    - Change in frequency of moderate to severe HF from baseline to Week 1 (all regions)
    - Change in PROMIS SD SF 8b total score from baseline to Week 12 (all regions).
    - Change in MENQOL total score from baseline to Week 12 (all regions).
- Population level summary:
    - Mean change in severity of moderate to severe HF from baseline to Week 4 (all regions except regulatory submission in the US).
    - Mean change in severity of moderate to severe HF from baseline to Week 12 (all regions except regulatory submission in the US).
    - Mean change in frequency of moderate to severe HF from baseline to Week 1 (all regions).
    - Mean change in PROMIS SD SF 8b total score from baseline to Week 12 (all regions).
    - Mean change in MENQOL total score from baseline to Week 12 (all regions).

  Treatment comparison will be based on differences in treatment group means for each endpoint.

1.2 Study Design

Figure 1–1: Study Schema



Approx. = approximately, ICF = Signing of informed consent form, Rand. = randomization

Multi-center, multi-country, double-blind, randomized, parallel-group, placebo-controlled, Phase 3 intervention study in post-menopausal women with vasomotor symptoms.

Total study duration for all participants: approximately 36 weeks (plus potential washout period), including

- Pre-screening/ Wash-out period (if applicable)

  After giving informed consent, but before starting formal screening procedures, participants will be withdrawn from prohibited concomitant medications.
- Screening: approximately 6 weeks
- Treatment: 26 weeks
- Follow-up: 4 weeks

Visit Frequency: Every 4-6 weeks

Participants will be randomized 1:1 to either
Elinzanetant 120 mg for 26 weeks
Placebo for 12 weeks, followed by elinzanetant 120 mg for 14 weeks
Randomization will be stratified by region: North America, rest of the world (ROW)
No interim analysis planned.
Primary analysis will be performed after database release after LPLV.

2. Statistical Hypotheses

The hypotheses for the primary and key secondary efficacy endpoints, are defined as:

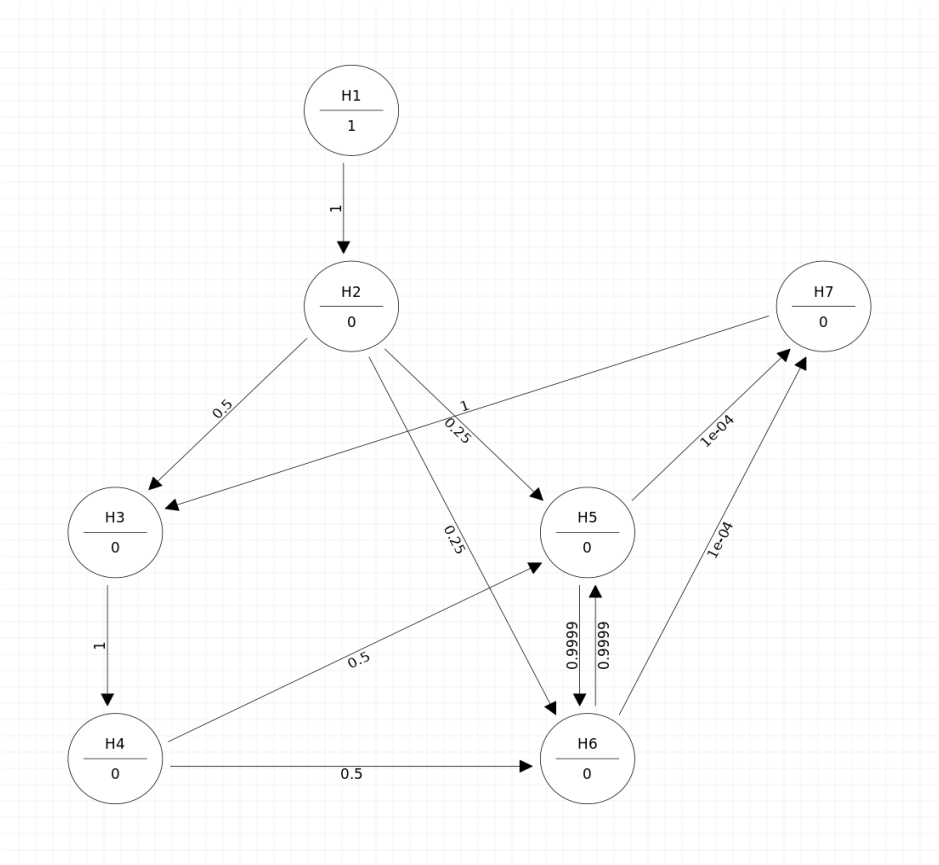
H1- $H_{01}: \mu_{1P} \leq \mu_{1V}$ versus $H_{11}: \mu_{1P} > \mu_{1V}$ where μ_{1P} and μ_{1V} stand for the mean change from baseline in the placebo (P) and verum (V) group in HF frequency at week 4.
H2- $H_{02}: \mu_{2P} \leq \mu_{2V}$ versus $H_{12}: \mu_{2P} > \mu_{2V}$ where μ_{2P} and μ_{2V} stand for the mean change from baseline in the placebo (P) and verum (V) group in HF frequency at week 12.
H3- $H_{03}: \mu_{3P} \leq \mu_{3V}$ versus $H_{13}: \mu_{3P} > \mu_{3V}$ where μ_{3P} and μ_{3V} stand for the mean change from baseline in the placebo (P) and verum (V) group in HF severity at week 4.
H4- $H_{04}: \mu_{4P} \leq \mu_{4V}$ versus $H_{14}: \mu_{4P} > \mu_{4V}$ where μ_{4P} and μ_{4V} stand for the mean change from baseline in the placebo (P) and verum (V) group in HF severity at week 12.
H5- $H_{05}: \mu_{5P} \leq \mu_{5V}$ versus $H_{15}: \mu_{5P} > \mu_{5V}$ where μ_{5P} and μ_{5V} stand for the mean change from baseline in the placebo (P) and verum (V) group in PROMIS SD SF 8b at week 12.
H6- $H_{06}: \mu_{6P} \leq \mu_{6V}$ versus $H_{16}: \mu_{6P} > \mu_{6V}$ where μ_{6P} and μ_{6V} stand for the mean change from baseline in the placebo (P) and verum (V) group in HF frequency at week 1.
H7- $H_{07}: \mu_{7P} \leq \mu_{7V}$ versus $H_{17}: \mu_{7P} > \mu_{7V}$ where μ_{7P} and μ_{7V} stand for the mean change from baseline in the placebo (P) and verum (V) group in MENQOL at week 12.

The data collected after Week 12 will be summarized in a descriptive manner.

A multiplicity adjustment strategy is defined in this trial using the graphical method (Bretz et al. 2009). The multiplicity adjustment strategy controls the overall Type I error rate at a one-sided $\alpha = 0.025$ level under any joint distribution of the test statistics corresponding to the above listed null hypotheses. The testing strategy for the seven null hypotheses is depicted in Figure 2–1.

The testing of the hypotheses will stop if the null hypothesis for either of the primary endpoints on the frequency of VMS at Week 4 or 12 (H_{01} & H_{02}) is not rejected. Given successful rejection of the first two primary endpoints at α, a weighted version of the Holm test will be applied to the significance level for the remaining hypotheses. H3 and H4 as the second important hypotheses are weighted as 0.5 in the testing strategy and will be sequentially tested in the second stage at $\alpha/2$. H5 & H6 were considered equally important and weighted at 0.25 each. H7 will be tested last only if H_{05} and H_{06} are successfully rejected. The level of significance for H5, H6, H7 depends on the outcome of the test for H3 & H4. The application of testing strategy is described in more details below.

Figure 2–1: Testing strategy for the seven hypotheses



The multiplicity adjustment strategy depicted in Figure 2–1 translates to the following serial testing algorithm: Let p_1 through p_7 denote the one-sided p-values corresponding to the hypotheses H_1 through H_7, respectively.

Step 1. Test the hypotheses H_{01} and H_{02} sequentially at the full α level, i.e., Reject H_{01} if $p_1 \leq \alpha$ and reject H_{02} if $p_1 \leq \alpha$ and $p_2 \leq \alpha$. Proceed to Step 2 if H_{02} is rejected.
Step 2. Test the hypotheses H_{03} and H_{04} sequentially at the $\alpha/2$ level, i.e., reject H_{03} if $p_3 \leq \alpha/2$ and reject H_{04} if $p_3 \leq \alpha/2$ and $p_4 \leq \alpha/2$. Proceed to Step 3.
Step 3. Test the hypotheses H_{05} and H_{06} using a weighted version of the Holm test. The following two cases need to be considered:

- H_4 is not rejected in Step 2:
  - **Step 3A**: Reject H_{05} if $p_5 \leq \alpha/4$ and proceed to Step 3B.
  - **Step 3B**: Reject H_{06} if $p_5 \leq \alpha/4$ and $p_6 \leq \alpha/2$ or if $p_5 > \alpha/4$ and $p_6 \leq \alpha/4$. Proceed to Step 3C if H_{05} is not rejected in Step 3A and H_{06} is rejected in Step 3B.
  - **Step 3C**: Reject H_{05} if $p_5 \leq \alpha/2$.
- H_4 is rejected in Step 2:
  - **Step 3A**: Reject H_{05} if $p_5 \leq \alpha/2$ and proceed to Step 3B.

- **Step 3B**: Reject H_{06} if $p_5 \leq \alpha/2$ and $p_6 \leq \alpha$ or if $p_5 > \alpha/2$ and $p_6 \leq \alpha/2$. Proceed to Step 3C if H_{05} is not rejected in Step 3A and H_{06} is rejected in Step 3B.
- **Step 3C**: Reject H_{05} if $p_5 \leq \alpha$.

Proceed to Step 4 if both H_{05} and H_{06} are rejected.
Step 4. Test the hypothesis H_{07} at the full α level if H_{04} is rejected in Step 2, i.e., reject H_{07} if $p_7 \leq \alpha$. Otherwise, if H_{04} is not rejected in Step 2 test H_{07} at the $\alpha/2$ level, i.e., reject H_{07} if $p_7 \leq \alpha/2$. Proceed to Step 5 if H_{07} is rejected.
Step 5. If H_{03} is not rejected in Step 2, re-test the hypothesis H_{03} at the α level, i.e., reject H_{03} if $p_3 \leq \alpha$. Proceed to Step 6 if H_{03} is rejected.
Step 6. If H_{04} is not rejected in Step 2, re-test the hypothesis H_{04} at the α level, i.e., reject H_{04} if $p_4 \leq \alpha$.

3. Sample Size Determination

A total of 370 participants (185 per arm) will be randomized in a 1:1 ratio to both arms. Assuming a drop-out rate of 10 % in the first 3 months, this will result in approximately 332 evaluable participants (166 per arm) who completed 12 weeks of treatment. The drop-out rate in months 4-6 is also expected to be 10 %, which will result in approximately 298 participants (149 per arm) who completed 6 months of treatment.

Non-evaluable participants are defined as participants that do not contribute data for the efficacy analysis, for example due to missing baseline information or no post-baseline data.

Assuming an overall screen failure rate of 65%, approximately 1058 participants need to be enrolled to achieve the required number of 370 randomized participants. More participants might be enrolled in case the screen failure rate is higher than anticipated. Additionally, further patients may be enrolled in case of an exceptionally high drop-out rate due to a trial-continuity issue.

A formal sample size estimation was performed for the efficacy analyses. The sample size was determined to power the study on the primary and key secondary endpoints listed in Section 1.1 (Table 1–1) and Section 2 at a minimum of 90%.

The sample size has been determined via simulation.

For each endpoint, the distribution of the effect for the placebo and treatment arms were built.

For endpoints 1, 2, 3, 4 & 6 (see the ordering in Section 2), we used the data from placebo arm in SWITCH-1 study (NCT03596762 2020) to build a distribution for the effect of placebo in the study. The placebo arm distribution then was shifted by the estimated treatment effect (from SWITCH-1) to represent the distribution of effects for the treatment arm.

For endpoints 5 & 7, there is limited data available in this population and therefore, a standard normal distribution (mean=0, std=1) was assumed for the placebo arm effect. A treatment effect of 0.4 is selected based on clinical team input.

A fixed level of correlation, 0.3, was assumed between the endpoints.

The assumed treatment effect and characteristics of the placebo distributions are presented in Table 3–1.

Table 3–1: Assumed treatment effect and characteristics of the placebo distributions

Endpoint	Treatment effect Treatment vs. placebo	Distribution of the placebo arm	Distribution parameters
1 – CFB HF Freq W4	-3.5	Normal	Mean= -2.29, Std= 3.632
2 – CFB HF Freq W12	-2	Normal	Mean= -4.43, Std= 4.29
3 – CFB HF Sev W4	-0.22	Mixture normal (Two normal distributions weighted equally at 0.5)	Mean1= -0.543, Std1= 0.562, Mean2= -0.105, Std2= 0.123
4 – CFB HF Sev W12	-0.26	Mixture normal (Two normal distributions weighted equally at 0.5)	Mean1= -0.897, Std1= 0.733, Mean2= -0.128, Std2= 0.185
5 – CFB PROMIS W12	-0.4	Standard Normal	Mean= 0, Std= 1
6 – CFB HF Freq W1	-2.36	Generalised normal	Psi= -0.43, Kappa= 0.628, Alpha= 2.277
7 – CFB MenQoL W12	-0.4	Standard Normal	Mean= 0, Std= 1

CFB = change from baseline, Freq. = Frequency; HF = hot flashes, MENQOL = Menopause-specific quality of life questionnaire, PROMIS SD SF 8b = Patient-Reported Outcomes Measurement

Information System Sleep Disturbance short-form 8b, Sev = Severity, Std = standard deviation, W = week.

Patient level data was simulated from the joint distributions of the endpoints. For each simulated study, the treatment effect was tested for the endpoints using a two sample T-test, where the assumption of normality is met, or a Wilcoxon sum-ranked otherwise. The power was calculated as the total number of rejected null hypotheses out of the number of simulated studies for each endpoint. The significance level (α) for the tests was adjusted based on the multiplicity strategy.

Under the assumptions given in Table 3–1, the study should be powered to achieve a statistical significance for all seven primary and key secondary endpoints at a minimum of 90% with 166 participants per arm.

4. Analysis Sets

For the purposes of analysis, the following analysis sets are defined in Table 4–1.

Table 4–1: Definition of the analysis sets

Analysis Set	Description
Enrolled	All participants who sign the informed consent form.
Full Analysis Set (FAS)	All randomized participants.
Safety Analysis Set (SAF)	All participants who receive at least one dose of study intervention.

Efficacy analyses will be based on the FAS and participants will be analyzed according to the randomized intervention. Safety analyses will be performed on the SAF and here participants will be analyzed according to the intervention received.

Documentation of validity findings and assignment of participants to analysis sets will be performed according to the sponsor's applicable Standard Operating Procedures.

Final decisions regarding the assignment of participants to analysis sets will be made during the review of study data and documented in the final list of important deviations, validity findings and assignment to analysis set(s).

Tables that are pre-specified to be displayed for FAS and SAF may be displayed for the FAS exclusively in case both analysis sets as well as the randomized intervention and the actual intervention received are identical.

5. Statistical Analyses

5.1 General Considerations

For a detailed description of the testing procedure including adjustments to the significance level, please see Section 2. The overall type I error rate will be controlled at 2.5 % one-sided. In general, confidence intervals (CI) will be two-sided with a confidence level of 95%.

The statistical evaluation will be performed by using the software SAS (release 9.4 or higher; SAS Institute Inc., Cary, NC, USA) and ValidR (version 3.5.2 or higher; Mango Solutions Ltd., UK).

All variables will be summarized by descriptive statistical methods. The number of data available, arithmetic mean, standard deviation, minimum, median, and maximum will be calculated for metric data. The geometric mean and geometric SD will be provided instead of the arithmetic mean and SD for variables where lognormal distributions are assumed. Frequency tables will be generated for categorical data. Where appropriate, the data will also be presented by visit or week, including the analysis of the changes from baseline.

Variables recorded in the electronic case report form (eCRF) and relevant derived variables will be shown in subject data listings, whereby only randomized patients will be included. Data from screening failures will only be shown in the 'Screening failure' listing.

Number of decimal places for summary statistics will be the following:

Table 5–1: Number of decimal places for summary statistics

Statistic	Number of digits
Minimum, maximum	Same as original data
Mean, median	1 more than in original data
SD	1 more than in original data
Frequencies (%)	1 digit
p-values	4 digits

5.1.1 Handling of Missing Data

This section describes the general handling of missing data. This includes data that would be meaningful for the analysis in terms of the study objectives but were not collected. The rules for handling the data that do not exist or are not considered meaningful for the analysis, because of an intercurrent event, are described in Section 5.3.2 and 5.3.4. No imputation of missing assessments will be performed for the secondary and exploratory endpoints unless it is specified otherwise.

All missing or partial data will be presented in the subject data listing as they are recorded on the electronic Case Report Form (eCRF) or electronic diary (eDiary).

For computation of durations, i.e., time between start and end dates of certain events and concomitant medication intake, a complete date is necessary.

The following rule will be applied to impute the missing start or end date of adverse events (AE)/ concomitant medications (CM):

An imputation range will be defined as the earliest possible date (lower range) and the latest possible date (upper range) of the occurrence of AE/CM. If AE/CM end date is available, this will be used as the latest possible AE/CM occurrence date in the imputation range.

If the active treatment start date falls within imputation range, incomplete AE/CM start date will be imputed as the active treatment start date. If the active treatment start date does not fall within the imputation range, and the placebo treatment start date falls within the imputation range, an incomplete AE/CM start date is imputed as the placebo treatment start date. Otherwise, the partially missing AE/CM start date will be imputed to the earliest date of the imputation range. Completely missing start date will not be imputed.

Partially missing AE/CM end date will be imputed as the latest possible date of AE/CM imputation range. Completely missing end date will not be imputed and presumed to be ongoing. The date of permanent discontinuation of study medication will be collected in eCRF. In case of missing date, it will be determined based on last medication intake date before the discontinuation using available eDiary data.

Unless specified otherwise, participants with missing baseline assessments will be excluded from all analyses that require the respective baseline assessment.

5.1.1.1 Electronic Participant-Reported Outcomes (ePROs)

There are two types of missing data for ePRO measures: missing data at the ‘form’ level and missing data at the ‘item’ level. Form level missing refers to a participant missing an entire PRO assessment for a given time point (e.g., a given day for the HFDD and a given week for the PROMIS SD SF 8b). In general, form level data may be missing due to participant’s early withdrawal from the study, inability to evaluate an endpoint at a particular time point, or non-compliance. By design of the eDiary, there will be no item level missing data for the respective ePRO questionnaires because the participants have to select an answer for an item in order to move on to the next item on the eDiary/electronic handheld device. This does not apply to the MENQOL (see below).

5.1.1.1.1 Hot Flash Daily Diary (HFDD)

Participants’ assessments of HF will be recorded electronically twice daily using the sponsor developed HFDD.

A diary day for the calculation of the frequency and severity of HF consists of the evening entry (Evening Hot Flash eDiary) and the morning entry (Morning Hot Flash eDiary) of the subsequent day. A day will be considered available for the calculation of the frequency and severity of HF, if at least the evening or the morning entry (of the subsequent day) is not missing. The daily number of hot flashes will be calculated as the sum of hot flashes documented in both the evening and morning diary. Retrospective data entry is possible for a limited time. For the morning diary these retrospective entries will be allowed on the same day between 11:00AM and 11:59PM. For the evening diary the retrospective entry option will be available between midnight until 10:59AM on the day after the missed entry. If only the evening or morning entry is available, then only this will be used for that particular day.

For the evaluation of the frequency and severity of HF, the daily HF assessments will be aggregated to a mean daily frequency and a mean daily severity HF score from the data of a particular week (see Section 5.3.1). In case data is missing for more than 2 days within a

week, the value for that particular week will be set to missing. Further details regarding imputation rules in case of missing week values are described in Section 5.3.2.

5.1.1.1.2 Patient-reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b (PROMIS SF 8b)

The PROMIS SF 8b will be filled out by the participants using the eDiary. All PROMIS SD SF 8b items must be answered to produce a valid total score. Handling of missing weekly assessments is described in Section 5.4.1.

5.1.1.1.3 Menopause Specific Quality of Life Scale (MENQOL)

The MENQOL will be filled out by the participants using the eDiary.

For the MENQOL, participants have the option to skip single questions (e.g. if they do not feel comfortable to complete them). Missing item score will be imputed, if the participant has responded to more than one half of the domain items (i.e. at least two items in the vasomotor domain, two items in the sexual domain, four items in the psychosocial domain and nine items in the physical domain of the MENQOL). The imputed score is calculated from values converted for analysis according to scoring table (Table 6–2) and rounded to a whole number. The imputed value is the mean for that item generated from the other subjects who responded as follows: For missing item score at baseline, all subjects who responded to the item are considered. For post-randomization, only the responses of the participants in the same treatment group are used. When data from other participants will be used for imputation half of the participants within a treatment group must have responded to the item before an imputed value can be calculated. If the participant answered ‘yes’ but did not indicate ‘how bothered’ she was (partially missing data), the value is imputed from her own answers by calculating the mean of her ‘bothered’ scores for all her ‘yes’ answers within that domain. If she answered ‘no’ to all other domain items, the imputed score would be generated from the mean of all the subjects who responded ‘yes’ to the same item as discussed above (Lewis, 2004). If missing item scores within a domain cannot be imputed (because participant responded to half or less than half of the domain items), the domain score will be set to missing and the total score will be calculated as the mean of the remaining available domains.

Further details of scoring and handling of missing weekly assessments are described in Section 6.2.3. and Section 5.4.1.

5.1.1.1.4 Beck Depression Inventory (BDI-II)

The BDI-II will be filled out by the participants using the eDiary. No imputation of missing assessments will be performed.

5.1.1.1.5 Insomnia Severity Index (ISI)

The ISI will be filled out by the participants using the eDiary. No imputation of missing assessments will be performed.

5.1.1.1.6 Patient Global Impression of Severity (PGI-S) and Change (PGI-C)

The PGI-S and PGI-C will be filled out by the participants using the eDiary at home. No imputation of missing assessments will be performed.

5.1.1.1.7 European Quality of Life 5-dimension 5-level questionnaire (EQ-5D-5L)

The EQ-5D-5L will be filled out by the participants using the eDiary. No imputation of missing assessments will be performed.

5.1.1.1.8 Sleepiness Scale

The Sleepiness Scale will be assessed together with the HFDD evening diary at bedtime during selected time periods. It consists of three items asking the participant to rate the degree of sleepiness felt in the morning (item 1), afternoon (item 2) and evening (item 3).

Missing item level responses will not be imputed, the average of item scores will be used to create a daily sleepiness score. A 7-day average score can be derived if a daily sleepiness score is available for at least 5 out of 7 of days, otherwise the 7-day average score will be set as missing. The baseline value will be calculated by aggregating all available days during the 14 days prior to start of treatment to a mean daily sleepiness. A minimum of 11 days should be available for the derivation of the baseline value.

5.1.2 Data Rules

Definition of baseline: The latest available valid measurement excluding unscheduled measurements at or before the start-of-treatment (for treated subjects) / randomization date (for randomized but not treated subjects). If the last observation available prior to randomization is the measurement from the Screening Visit, this would be used as the baseline value. In case a wash-out period was necessary, measurements taken during the pre-screening visit will not be used as a baseline value unless specified otherwise.

Change from baseline: To calculate the absolute change from baseline the baseline value will be subtracted from the value under treatment/follow-up, i.e.,

Absolute change = post baseline value – baseline value.

Some parameters will be additionally analyzed as relative change defined as

Relative change = 100 * [(post baseline value – baseline value) / baseline value].

Laboratory values <X or >Y: For laboratory values lower than a limit of detection X (reported as <X), half the value of X will be used for analysis (i.e., for values reported as "<X", "<X.X"," <X.XX" or etc. the value for analysis will be derived by "X/2", "X.X/2"," X.XX/2" etc). Difference between 2 values <X will be assigned a value of 0. Ratio between 2 values <X will be assigned a value of 1. For values which are higher than a limit of detection Y (reported as, >Y), the value of Y will be used for analysis.

Repeated measurements at the same visit after start of treatment: If more than one post-randomization measurement is available for a given visit, the first available observation will be used in the data summaries and all observations will be presented in the data listings.

Start of treatment refers to the first drug intake and will be identified based on the eCRF entry.

Time to event: The randomization date will be used as the start date for the calculation of time to event. Therefore, the time of event will be defined as "date of event – randomization date +1".

Unscheduled Assessments: Extra assessments (e.g. laboratory data or vital signs) associated with non-protocol visits will be included in listings, but not in the summary tables or figures, unless specified otherwise.

End-of-treatment ePRO assessments after premature treatment discontinuation:

Participant-Reported Outcome assessments (PROMIS SD SF 8b, MenQoL, EQ-5D-5L, BDI-II, ISI, PGI-C, and PGI-S) are scheduled to occur during the following days:

Table 5–2: Scheduled assessment weeks for ePROs

Questionnaires	Assessment Week	Completion days (inclusive)
PROMIS SD SF 8b, PGI-S, EQ-5D-5L	1	8-9
PROMIS SD SF 8b, PGI-S, EQ-5D-5L	2	15-16
PROMIS SD SF 8b, PGI-S, EQ-5D-5L	3	22-23
PROMIS SD SF 8b, PGI-S, EQ-5D-5L, MenQoL, ISI, PGI-C, BDI-II	4	29-30
PROMIS SD SF 8b, PGI-S, EQ-5D-5L, MenQoL, ISI, PGI-C, BDI-II	8	55-57
PROMIS SD SF 8b, PGI-S, EQ-5D-5L, MenQoL, ISI, PGI-C, BDI-II	12	83-85

First, scheduled assessments during the completion days will be assigned to the respective week.

Only in case no scheduled assessment is available for a particular week, assessments done at EoT, unscheduled timepoints or follow-up (in this order of priority) will be used, and they will be mapped to the next available, protocol-planned assessment week as described below.

- If the assessment occurs during the completion days as above, it will be assigned to the respective week.
- Up to and including Week 3 (only PROMIS SD SF 8b, EQ-5D-5L, PGI-S): If the EoT/ follow-up/ unscheduled assessment is on 5th day of the week or later, then the assessment will be mapped to the respective week (i.e., assessments on Days 6-7* will be mapped to Week 1, Days 12-14 to Week 2 and Days 19-21 to Week 3). Otherwise, no mapping will occur, and the respective week will be considered missing. If multiple assessments are available within a particular week, the last assessment (i.e., the one closest to the scheduled assessment) will be used for mapping.
- From Week 4 onwards: a 2-week window will be used to map the available assessments to Week 4 (Days 15-28), Week 8 (Days 43-54*), or Week 12 (71-82*). A similar 2-week window will be used for Week 16 and Week 26. Within each 2-week-window, the last available assessment (i.e., the one closest to the scheduled assessment) will be used for mapping.

  *Week 1 = Days 2-8; Day 8 (week 1), Days 55/56 (week 8) and Days 83/84 (week 12) are scheduled assessments, and no mapping would be required, therefore, these days are not considered in the window for mapping.
- In case an assessment is assigned to Week 2 or 3 and to Week 4, based on the above rules, it will be shown for Week 2 or 3 in the descriptive tables. The Week 4 assessment will only be used in the model-based analyses.

- Assessments from follow-up visits will be shown as follow-up in the descriptive tables, and the value mapped to a specific treatment week will only be used in the model-based analyses.
- In the descriptive tables, data collected after premature discontinuation of treatment up to week 12 will not be shown for the respective treatment week. Assessments occurring up to 2 days after discontinuation (inclusive) will still be considered under treatment and shown for the respective week. Out of the assessment collected post treatment discontinuation (+ 2 days), the one closest to the planned time point of follow-up (i.e. discontinuation date + 28 days) will be shown as follow-up assessment. In case two assessments are equally close to that date, the first one will be selected.

Definition of phases: The two phases in this study represent the approximately first 12 weeks of the study, in which participants receive either placebo or elinzanetant 120mg (this is referred as drug part 1) and the approximately 14 weeks when all participants receive elinzanetant (this is referred as drug part 2). The start and end date of the two phases of treatment is defined as following:

- Phase 1:
  - Start date = randomization
    (*as the first dose will be taken on site, this should match the data in Rave*).
  - End date = first day of drug part 2 intake – 1 day based on the eCRF entry)
    (*this is for participants who stayed in the study for at least 12 weeks and received the drug part 2*);

    OR

    (For participants who have not received the drug part 2 due to early discontinuation of randomized treatment or study withdrawal)

    [Week 12 assessment date or the post treatment phase end date, whichever is earlier
    (*this is for participants who discontinued from study drug but agreed to complete the assessments until Week 12*)
    or
    Study drug discontinuation date or EOT assessment date, whichever is later
    (*this is for participants who discontinued from randomized drug and decided to go to the follow up period or withdraw from the study*)].

Note: HFDD morning diary and PRO questionnaires which were completed one day after the phase 1 end date will be assigned to phase 1. For participants who were randomized but not treated, the end data = randomization date.

- Phase 2:
  - Start date = first day of drug part 2 intake (identified based on the eCRF entry)
  - End date = end of treatment period date
    (*For participants who completed the treatment period*)
    or
    Study drug discontinuation date or EOT assessment date, whichever is later
    (*For participants who discontinue from treatment/study prematurely during the second phase*).

Derivation of intercurrent event (ICE): Intercurrent events occurred from randomization day to day 84 inclusive will be flagged.

For the "permanent discontinuation of randomized drug" and "intake of prohibited medication having impact on efficacy", if the ICE occurs on or before the 5th day of the week, it will be flagged for the respective week. If the ICE occurs on the 6th or 7th day of the week, it will be flagged from the following week. For example, if the participant permanently discontinues from the randomized drug on the 2nd day of week 4, the ICE will be flagged from week 4 onwards. If the discontinuation from the randomized drug occurs on the 6th day of week 4, then the ICE will be flagged from Week 5 onwards. For the "intake of prohibited medication having impact on efficacy" ICE, if the end date for the impact on efficacy occurs on the 1st or 2nd day of the week, the respective week should not be flagged. If the end date of impact on efficacy occurs on and after the 3rd day of the week, the week should be flagged.

The "permanent discontinuation of randomized drug" ICE will be flagged for participants who were randomized but not treated. These participants have an event at randomization for the Kaplan-Meier analysis on the time from randomization to the first occurrence of the intercurrent event "Permanent discontinuation of randomized treatment". For the main analytical approach, the missing post-baseline data will be imputed as described in Section 5.3.2.

Derivation of follow up weeks for HFDD: In the descriptive tables, data collected after premature discontinuation of treatment up to week 12 will not be shown for the respective treatment week, but as follow-up data. The follow-up weeks are counted from last treatment date +1 onwards. Post-treatment days of the week in which the discontinuation of randomized drug occurs will be included in the last treatment week and the first follow up week. For example, if a participant's last dose intake was on day 39 (4th day of week 6), week 6 will include data from days 36-42 and the follow-up week 1 will include the data from days 40-46. For subjects who prematurely discontinued from study drug but agreed to continue with the scheduled visits, only 4 weeks of follow up will be counted.

5.2 Participant Dispositions

The number of participants enrolled, randomized and valid for the FAS and SAF, will be summarized overall and by treatment group, region, country and study site. A listing of the participants' assignment to the FAS and SAF, and the reasons for exclusion will also be provided by treatment group.

Other summary statistics will include:

- number of screening failures and the reason for failed screening
- number of participants randomized but not treated (overall and by treatment group).
- number of participants who completed the study, i.e., completed all phases (screening, treatment and follow up) of the study including the last visit.

A disposition summary for each study period (screening, treatment and follow-up period, see Section 1.2) will be presented summarizing the number of participants starting and completing the respective study period, the number of participants discontinuing it and the primary reason for discontinuation. This will also be presented overall and for each treatment group. The table will include COVID-19 pandemic associated reasons for discontinuation, i.e., the information whether decision for discontinuation was due to participant specific reasons or due to other reasons.

In addition, the number of participants with important deviations and validity findings will be presented overall and for each treatment group. The frequencies of each important deviation and validity finding will be presented by treatment group and in total.

Additionally, Kaplan-Meier plots for the time from randomization to the first occurrence of the intercurrent events “Permanent discontinuation of randomized treatment” and “Intake of prohibited concomitant medication having impact on efficacy” will be provided by treatment group, where applicable. If “Permanent discontinuation of randomized treatment” did not occur by day 84, the participant will be censored at week 12. If “Intake of prohibited concomitant medication having impact on efficacy” did not occur by day 84, the participant will be censored at week 12 or at the time of dropping out of the study, whichever occurs earlier.

The number of observed intercurrent events will be summarized by week and by treatment group and overall.

5.3 Primary Endpoints Analysis

The primary endpoints in this study are listed below.

For regulatory submission in the US:

- Mean change in frequency of moderate to severe HF from baseline to Week 4 (assessed by HFDD).
- Mean change in frequency of moderate to severe HF from baseline to Week 12 (assessed by HFDD).
- Mean change in severity of moderate to severe HF from baseline to Week 4 (assessed by HFDD).
- Mean change in severity of moderate to severe HF from baseline to Week 12 (assessed by HFDD).

For all other regions except regulatory submission in the US:

- Mean change in frequency of moderate to severe HF from baseline to Week 4 (assessed by HFDD).
- Mean change in frequency of moderate to severe HF from baseline to Week 12 (assessed by HFDD).

5.3.1 Definition of Endpoints

Participants’ assessments of HF will be recorded electronically twice daily using the Sponsor developed Hot Flash Daily Diary (HFDD). See details in Section 6.2.1.

Baseline value of frequency of moderate to severe HF:

The baseline value will be calculated by aggregating all available days during the 14 days prior to start of treatment to a mean daily frequency as (total number of moderate to severe HF during the 14 days prior to start of treatment) / (total number of available days with data). As per inclusion criteria 9 (see protocol), at least 11 days will be available for the derivation of the baseline value.

Frequency of moderate to severe HF during treatment:

The frequency of moderate to severe HF for each week during the treatment period will be calculated using the available data during that particular week. Specifically, for Week 1 Days 2-8 will be used, for Week 4 Days 22-28 will be used and for Week 12 Days 78-84 will be used (Day 1 corresponds to start of treatment). These data will be aggregated to a mean daily frequency as (total number of moderate to severe HF during that week) / (total number of available days with data during that week). In case data is missing for more than 2 days within a week, the value for that particular week will be set to missing.

Mean change in frequency of moderate to severe HF from baseline to Week 4 and Week 12:

This will be calculated as the difference in the mean daily frequency of moderate to severe HF at Week 4 or Week 12, respectively, from the baseline value of the mean daily frequency of moderate to severe HF.

Baseline value of severity of moderate to severe HF:

The mean daily severity during baseline will be calculated for the available days as [(2 x number of moderate HF) + (3 x number of severe HF)] / (total number of moderate to severe hot flashes on that day). When no moderate or severe HF are reported for a particular day, the mean severity for that day will be set to 0. The baseline value will be calculated by averaging the mean daily severity of the available days during the 14 days prior to start of treatment. As per inclusion criteria 9 (see protocol), at least 11 days will be available for the derivation of the baseline value.

Severity of HF during treatment:

The mean daily severity during treatment will be calculated for the available days as [(1 x number of mild HF) + (2 x number of moderate HF) + (3 x number of severe HF)] / (total number of mild, moderate and severe hot flashes on that day). When no HF are reported for a particular day, the mean severity for that day will be set to 0.

Similar to the frequency, the severity of HF for each week during the treatment period will be calculated using the available data during that particular week. Specifically, for Week 4 Days 22-28 will be used and for Week 12 Days 78-84 will be used (Day 1 corresponds to start of treatment). To obtain the post-baseline severity of HF during a particular week, the weekly data will be aggregated by averaging the mean daily severity of HF of the available days during that week. In case data is missing for more than 2 days within a week, the value for that particular week will be set to missing.

Mean change in severity of moderate to severe HF from baseline to Week 4 and Week 12:

This will be calculated as the difference in the severity of HF at Week 4 or Week 12, respectively, from the baseline value of the severity of moderate to severe HF.

5.3.2 Main Analytical Approach

The main estimand to answer the clinical question of the study is described in Section 1.1.

The frequency and severity of moderate to severe HF and the change from baseline as well as the relative change (%) from baseline will be summarized using descriptive statistics (see Section 5.1) by treatment group and by week. The change from baseline over time in the

frequency of moderate to severe HFs will be shown using line plots for means together with 95% CIs by treatment group.

Inference about the primary endpoints will be done based on a mixed model for repeated measures (MMRM). Prior to performing modeling with the MMRM, missing data will be handled in alignment with the estimand strategies for ICEs (described in Section 1.1).

Table 5–3 provides a detailed overview of the MMRM model for the change from baseline in frequency of HF. A similar analysis model will be conducted for the change from baseline in severity of HF endpoints.

Table 5–3: Model description

Dependent variable:	Change from baseline in frequency of moderate to severe HF at Week 1, Week 4, Week 8 and Week 12
Covariates (continuous variables):	Baseline mean daily frequency of HF
Factors (class variables):	Treatment, Week, Region (stratification factor)
Interaction terms:	Baseline*Week Treatment*Week
Covariance structure:	Unstructured (UN) If the convergence cannot be attained with the unstructured correlation matrix, the following alternative structures will be attempted in the specified order: autoregressive(1) (AR(1)), and compound symmetry (CS). The first structure in this list with which the model converges will be used. If AR(1) structure is used, a random subject intercept will also be included in the model. This will need to be applied for each imputed dataset.
Equation	$Y_{ijkl} = \mu + \beta x_i + t_k + r_l + v_j + (tv)_{jk} + \alpha(xv)_{ij} + s_i + \varepsilon_{ijkl}$ where Y_{ijkl} is the change from baseline in frequency of HF to Week j for subject i (with treatment k and region l); μ is the intercept, β is the baseline covariate effect (baseline coefficient), x_i is the baseline frequency HF for subject i, t_k is the fixed effect of treatment k *(k=Elinzanetant, placebo)*, r_l is the fixed effect of region l *(l=North America, ROW)*, v_j is the fixed effect of Week j *(j=1, 4, 8 & 12),* $(tv)_{jk}$ is the interaction effect of treatment k by Week j, α is the covariate effect for the interaction (xv) $(xv)_{ij}$ is the interaction effect of baseline frequency HF of subject i by Week j, $s_i \sim$ Normal $(0, \sigma_s^2)$ is the random effect of subject i *(only if AR(1) covariance structure is used),*

	$\varepsilon_{i\,jkl} \sim$ Normal $(0, \sigma^2)$ represents the residual variance component with corr $(\varepsilon_{i\,j}, \varepsilon_{i\,j'}) = \rho_{jj'}, j \neq j'$.

SAS code:	`PROC MIXED data=DATA;`
	
	`CLASS subject treatment region Week;`
	`MODEL change = baseline treatment region Week treatment*Week baseline*Week /ddfm=KR outp=resid s;`
	`REPEATED Week / subject=subject type=un;`
	`RANDOM subject; /* only if AR(1) covariance structure is used */`
	
	`LSMEANS treatment*Week /cl diff;`
	
	`ESTIMATE 'Elinzanetant - Placebo at Week 1' treatment 1 -1 treatment*Week 1 0 0 0 -1 0 0 0 /cl;`
	`ESTIMATE 'Elinzanetant - Placebo at Week 4' treatment 1 -1 treatment*Week 0 1 0 0 0 -1 0 0 /cl;`
	`ESTIMATE 'Elinzanetant - Placebo at Week 12' treatment 1 -1 treatment*Week 0 0 0 1 0 0 0 -1 /cl;`
	
	`ODS OUTPUT TESTS3=TYPE3_eff LSMeans=LSMEAN ESTIMATES=ESTIM;`
	`RUN;`

A treatment policy strategy will be applied to handle all of the specified ICEs in the main estimand. According to this strategy, all collected data should be utilized in the analysis irrespective of occurrence of the ICEs.

Although all study participants are expected to be followed after ICEs, some missing data may occur. Missing values that occur while participants continue on their randomized treatment and simply represent missed assessments will be assumed missing at random (MAR). Such missing values are likely to be intermittent and will be imputed using a Monte Carlo Markov Chain (MCMC) multiple imputation (MI) method for imputation of non-monotone missing data. For the baseline value to be imputed, at least 2 post-baseline values out of the 4 post-baseline values (i.e. Week 1, 4, 8 & 12) need to be available. Possible missing baseline values will be imputed together with other intermittent missing values. Absolute values will be used for imputation steps.

Missing values that occur after the discontinuation of randomized treatment will be imputed using a MI model (Rubin, 1987) such that it aligns with the treatment policy estimand. The MI regression model will include auxiliary variables indicating (yes/no) whether the participant continued on randomized treatment at each visit (i.e., Week 1, 4, 8 & 12). These variables will be included in the imputation model to account for the occurrence of permanent discontinuation of randomized treatment (Guizzaroet al., 2021). In addition to the indicator variables, baseline HF, region and treatment as specified in Table 5–3 as well as previous post-baseline endpoint values will also be included in the imputation model. In case there are less than 10 participants (who permanently discontinued the randomized treatment) across both treatment groups with available post-discontinuation data at any intermediate time point, the indicator variable will not be included in the imputation model for that time point.

After the completion of imputation step, change from baseline values will be calculated and each imputed dataset will be analyzed using the MMRM model specified in Table 5–3 to obtain an estimate of the mean change from baseline and of the treatment effect. The results will be combined using the Rubin's rule (Rubin, 1987) to obtain an overall estimated mean

change from baseline and an overall estimated treatment effect. A total of 500 multiple imputation steps will be used.

Details regarding how the ICEs will be identified and specifics regarding the reasons for the occurrence of ICE are described in Section 6.5. All the ICEs will be reviewed prior to study unblinding.

The results from the main analysis will be presented with the number of participants, estimated mean change from baseline and standard error (SE) for each treatment arm at Weeks 1, 4 and 12 as well as the estimated treatment difference (elinzanetant - placebo), SE of the difference, associated 95% CI and P-value (one-sided). A plot of the model-based estimates with SE will be provided by both treatment groups.

5.3.3 Sensitivity Analyses

5.3.3.1 Assessment of normality assumption

The assumption for normality in the main analysis (see Section 5.3.2) will be evaluated by graphical tools (i.e., qqplot and plot of residuals against predicted values). This will be assessed based on the observed data before multiple imputation is applied.

For the case of extreme violations of the normality assumption, a non-parametric rank ANCOVA will be carried out as sensitivity analysis and the Hodges-Lehmann estimate will be calculated as estimate of the treatment effect (see Stokes et al. 2012). This will be carried out separately for the following primary/key secondary endpoints, change from baseline of HF frequency at Week 4, HF frequency at Week 12, HF severity at Week 4 and HF severity at Week 12. Specifically, a Cochran-Mantel-Haenszel score test will be applied to the residuals of a regression model on rank-transformed data while adjusting for baseline and the stratification factor region.

The methodology described in Stokes et al (2012) will be applied as follows using the SAS pseudo code given below. First, the values of the change from baseline variable as well as baseline covariate will be transformed to standardized ranks by region, using fractional ranks and mean method for ties:

```
proc rank data=datain nplus1 ties=mean out=ranks;
                by region;
                var baseline change_from_baseline;
run;
```

Afterwards separate regression models will be fitted within each region using the standardized rank values of the change from baseline and the baseline as dependent and independent variable, respectively. Residuals from these regression models will be captured for further testing of differences between treatment groups:

```
proc reg data=ranks;
                by region;
                model change_from_baseline = baseline;
                output out=residuals r=resid;
run;
```

Finally, the stratified Cochran-Mantel-Haenszel (CMH) test using the values of the residuals as scores will be used to compare the two treatment groups:

```
proc freq data=residuals;
                tables region*treatment*resid / CMH2;
                ods output cmh=cmhstat;
run;
```

The non-parametric rank ANCOVA described above will be applied to each imputation dataset within the multiple imputation procedure described earlier. Before combining the results of the CMH test using Rubin's rule, a normalizing transformation using the Wilson-Hilferty transformation as described in Ratitch , Lipkovich et al. (2013) will be applied. Let $cmh^{(m)}$ be the CMH statistic computed for the mth imputed dataset and *df* the corresponding degrees of freedom associated with the CMH statistic. The standardized test statistic for the mth imputation $st\_cmh^{(m)}$ is then given by

$$st\_cmh^{(m)} = \frac{\sqrt[3]{\frac{cmh^{(m)}}{df}} - \left(1 - \frac{2}{9 \times df}\right)}{\sqrt{\frac{2}{9 \times df}}}$$

The corresponding standard error is equal to 1. This standardized statistic together with its standard error will be passed on to PROC MIANALYZE to perform a combined CMH test using Rubin's rule.

In addition to the assessment of the treatment effect based on the CMH statistic as described above, the Hodges-Lehmann estimate of the median difference will be provided as estimate of the magnitude of the unadjusted treatment effect. The Hodges-Lehman estimate together with its asymptotic standard error will be computed for each imputation dataset and subsequently Rubin's rule will be used to combine the results. The Hodges-Lehmann estimate can be obtained using the following pseudo-SAS code:

```
proc npar1way hl;
                class treatment;
                var change_from_baseline;
                output out=HL;
run;
```

The two-sided p-value based on the non-parametric rank ANCOVA, as well as the Hodges-Lehmann based estimate of the size of the treatment effect together with the corresponding standard error and 95 % confidence interval will be reported. To assess statistical significance the p-value from the non-parametric rank ANCOVA will be used.

5.3.3.2 Tipping point analysis

A tipping point analysis will be applied to assess the sensitivity of the main analysis results to modeling of the missing data that occur in presence of ICEs as described in Section 5.3.2. This will be done by applying an unfavorable additive shift (referred to as delta adjustment) to the values imputed by the MI model for the main analysis in the elinzanetant arm. For the endpoints related to the frequency of HF, the adjustments will be applied with delta values of 1, 2, 3, 4, etc. in each successive tipping point iteration until a tipping point is attained. For the endpoints related to severity of hot flashes, the adjustments will be applied with delta

values of 0.2, 0.4, 0.6, etc. in each successive tipping point iteration until a tipping point is attained. Additional details on the tipping point sensitivity analysis are provided in Section 6.4.

5.3.4 Supplementary Analyses

Two supplementary estimands are defined for this trial:
The hypothetical strategy will be used for handling temporary treatment interruption and permanent treatment discontinuation due to COVID-19, administrative and any other treatment unrelated reasons in both supplementary estimands. Intake of prohibited medications and permanent treatment discontinuation with initiation of alternative VMS treatment will be handled by the hypothetical strategy in the first supplementary estimand and by the composite strategy, assuming the treatment failure, in the second supplementary estimand.

5.3.4.1 First Supplementary Analysis

The information about ICEs and strategies to address them for the first supplementary estimand is provided in Table 5–4. The other estimand attributes are the same as for the main estimand (see Section 1.1).

Table 5–4: First Supplementary Estimand: ICEs and Strategies to address them

ICEs**	Reason for ICE	Strategy	Data handling method
Temporary Treatment interruption*	AEs (treatment related/unrelated)	**Treatment policy**	Utilise the collected data after ICE.
	COVID-19 and administrative reasons	**Hypothetical**, If participant complied with treatment.	Model outcomes during the ICE under the hypothetical scenario.
Permanent discontinuation of randomized treatment	AEs (treatment related/unrelated) or Lack of efficacy		
	For participants who remained untreated/on background therapy.	**Treatment policy**	Utilise the collected data after ICE.
	For participants who initiate alternative VMS treatment	**Hypothetical**, If participant remained untreated.	Model outcomes after the ICE under the hypothetical scenario.
	Other treatment-unrelated reasons, including COVID-19	**Hypothetical,** If participant did not discontinue the randomized treatment.	Model outcomes after the ICE under the hypothetical scenario.
Intake of prohibited concomitant	All reasons	**Hypothetical,** If participant did not take prohibited concomitant	Model outcomes under the hypothetical scenario during the period of

medication having impact on efficacy	medication.	confounding.

ICE=Intercurrent event, AE=Adverse event, VMS = Vasomotor symptoms
*Definition of temporary treatment interruption:
Week 1 = Treatment taken on <5/7 days during week 1.
Week 4 = Treatment taken <80% during weeks 1-4 OR treatment taken on <5/7 days during either week 3 or 4.
Week 8 = Treatment taken <80% during weeks 1-8 OR treatment taken on <5/7 days during either week 7 or 8.
Week 12= Treatment taken <80% during weeks 1-12 OR treatment taken on <5/7 days during either week 11 or 12.
**ICEs will be reviewed prior to the study unblinding.

The details of data handling and missing data imputation for the first supplementary analysis of the primary endpoints are summarized in the Table 5–5. A pattern-mixture model using multiple imputation (MI) will be used to impute missing or discarded values that occur in presence of ICEs in a way that aligns with the ICE strategies.

The rows of Table 5–5 represent various patterns (subsets) of participants with a specific type of an ICE. For participants from each pattern who have missing data during the time frame of the ICE (Please see Sections 6.5.3 & 6.5.1 for the description of time frame for intake of prohibited medication and temporary treatment interruption ICEs. The time frame for the permanent treatment discontinuation is defined as the time after discontinuation) or unobservable data under the assumed hypothetical scenario, a reference group is specified in the last column of the table.

Available data of participants from the corresponding reference group will be used to model the distribution of missing and unobservable data in each pattern. For the baseline value to be imputed, at least 2 post-baseline values out of the 4 post-baseline values (i.e., Week 1, 4, 8 & 12) need to be available. In case there are less than 10 participants with available post-ICE data in the defined reference group, at any intermediate time point, the missing data for that pattern will not be imputed at that time point. Additional details are provided in Section 6.3.

Table 5–5: First supplementary analysis data handling and missing data imputation strategy.

Pattern ID	ICE and the associated reason	ICE strategy and data handling	Reference group for modeling the distribution of missing or unobservable data
1	Intake of prohibited medications Any reason	**Hypothetical strategy** If participant did not take prohibited concomitant medication. Discard the collected data after the intake of prohibited medication(s) for a certain time frame based on the medication class (See Section 6.5.3 for details).	Participants, from the same arm, who comply with treatment regimen** and have observed data at the time point that requires imputation.
	Temporary treatment interruption*	**Hypothetical strategy** If participant complied with	

Pattern ID	ICE and the associated reason	ICE strategy and data handling	Reference group for modeling the distribution of missing or unobservable data
	Due to COVID-19/ administrative or any other treatment unrelated reasons	treatment. Data collected for a specific assessment during/after a temporary treatment interruption will be discarded for that assessment Week only.	
2	Temporary treatment interruption* Due to AE	**Treatment policy strategy** Utilize all the collected data in the analysis.	Participants, from the same arm, who have available data before and during the treatment interruption. MI will be used to model change from a time point before treatment interruption to a time point with treatment interruption (Copy Increment from Reference imputation strategy).
3	Permanent treatment discontinuation Due to AE/lack of efficacy	**Treatment policy/Hypothetical strategy** If participant remained untreated/on background therapy after permanent discontinuation of the randomized treatment, use the treatment policy strategy and utilize the collected data. Otherwise, if participant initiated an alternative treatment, use the hypothetical strategy under the hypothetical scenario "If participant did not initiate an alternative treatment after permanent discontinuation of randomized treatment". Discard data after the ICE.	Missing values that occur after the ICE will be imputed using a similar MI model as used in the main analysis of the primary endpoint, i.e., the MI regression model will include auxiliary variables indicating (yes/no) whether the participant continued on randomized treatment at each visit (Guizzaro et al., 2021).
4	Permanent treatment discontinuation Due to COVID-19/ administrative or any other treatment unrelated reasons	**Hypothetical strategy** If participant did not discontinue the randomized treatment due to the specified reason. Any data collected after permanent treatment discontinuation will be discarded.	All other participants from the same arm (i.e., participants without a similar ICE at the time point that requires imputation).

*If more than one interruption for different reasons is recorded during the specified weeks (please see the definition of ICE in Section1.1), the data handling rule and missing data imputation will be performed based on the reason in the following pre-specified order: 1) AE, 2) Due to COVID-19/

administrative or any other treatment unrelated reasons.
** Complying with treatment regimen is defined as having no ICEs (i.e., no early discontinuation of randomized treatment, no treatment interruption in the respective week (definition in Table 1–2) and no intake of prohibited concomitant medications.

Missing values that occur in the time frame outside of any identified ICE and simply represent missed assessments will be assumed missing at random (MAR). Such missing values are likely to be intermittent and will be imputed using a Monte Carlo Markov Chain (MCMC) MI method for imputation of non-monotone missing data. Additional details are provided in Section 6.3.

A pattern-mixture model using multiple imputation will be applied to impute missing and unobservable data to mimic the distribution of observed data in the appropriate reference group as described in Table 5–5, conditional on covariates and partially observed data. The imputations will be done sequentially in patterns 1 through 4. The reference group for imputation in patterns 3 and 4 may contain participants who have some values imputed during the imputation of preceding patterns. Consequently, the MI model for pattern 4 will reflect a mixture of participants, some of which fully comply with the randomized treatment and some with other types of ICEs. Additional details of the pattern-mixture modeling with multiple imputation are provided in Section 6.3.

Each imputed dataset will be analyzed using the MMRM model specified in Table 5–3 to obtain an estimate of the treatment effect. The results will be combined using Rubin's rule (Rubin, 1987) to obtain an overall estimated treatment effect.

The same covariables will be included in the imputation and the analysis step as specified in Section 5.3.2.

The results from the supplementary analysis will be presented in a similar way as to the main analysis (see Section 5.3.2.).

5.3.4.2 Second Supplementary Analyses

The details of the second supplementary estimand can be found in Table 5–6 and Table 5–7. Similar analysis steps (i.e., imputation, modeling and combining results) as for the first supplementary analysis will be applied to the secondary supplementary analysis (see Section 5.3.4.1).

For the intake of prohibited medication and permanent treatment discontinuation after which participants initiated alternative VMS treatments, the composite strategy will be used to handle the data. The attributes for this estimand are the same as the main estimand except for following:

- Variable: Efficacy will be assessed based on 2 or 4 primary endpoints depending on the region (regulatory submission in the US only/all regions) as listed below:
    - Change in frequency of moderate to severe HF from baseline to Week 4 in the absence of ICEs (i.e., intake of prohibited medication or initiation of alternative VSM treatment after permanent discontinuation of randomized treatment) or no change from baseline otherwise (all regions).
    - Change in frequency of moderate to severe HF from baseline to Week 12 in the absence of ICEs (i.e., intake of prohibited medication or initiation of alternative VSM treatment after permanent discontinuation of randomized treatment) or no change from baseline otherwise (all regions).

- Change in severity of moderate to severe HF from baseline to Week 4 in the absence of ICEs (i.e., intake of prohibited medication or initiation of alternative VSM treatment after permanent discontinuation of randomized treatment) or no change from baseline otherwise (regulatory submission in the US only).
- Change in severity of moderate to severe HF from baseline to Week 12 in the absence of ICEs (i.e., intake of prohibited medication or initiation of alternative VSM treatment after permanent discontinuation of randomized treatment) or no change from baseline otherwise (regulatory submission in the US only).

- Intercurrent Events (ICEs): see Table 5–6.

Table 5–6: Second Supplementary Estimand: ICEs and Strategies to address them

ICEs**	Reason for ICE	Strategy	Data handling method
Temporary Treatment interruption*	AEs (treatment related/unrelated)	**Treatment policy**	Utilise the collected data after ICE.
	COVID-19 and administrative reasons	**Hypothetical** If participant complied with treatment	Model outcomes during the ICE under the hypothetical scenario.
Permanent discontinuation of randomized treatment	AEs (treatment related/unrelated) or Lack of efficacy		
	For participants who remained untreated/on background therapy.	**Treatment policy**	Utilise the collected data after ICE.
	For participants who initiate alternative VMS treatment	**Composite** Treatment failure (i.e., no change from baseline)	Model outcomes after the ICE under the treatment failure.
	Other treatment-unrelated reasons, including COVID-19	**Hypothetical** If participant did not discontinue the randomized treatment.	Model outcomes after the ICE under the hypothetical scenario.
Intake of prohibited concomitant medication having impact on efficacy	All reasons	**Composite** Treatment failure (i.e., no change from baseline)	Model outcomes under the treatment failure during the period of confounding.

*Definition of temporary treatment interruption:
Week 1 = Treatment taken on <5/7 days during week 1.
Week 4 = Treatment taken <80% during weeks 1-4 OR treatment taken on <5/7 days during either week 3 or 4.
Week 8 = Treatment taken <80% during weeks 1-8 OR treatment taken on <5/7 days during either week 7 or 8.
Week 12= Treatment taken <80% during weeks 1-12 OR treatment taken on <5/7 days during either

week 11 or 12.
**ICEs will be reviewed prior to the study unblinding.
ICE=Intercurrent event, AE=Adverse event, VMS = Vasomotor symptoms

Table 5–7: Second supplementary estimand data handling and missing data imputation strategy

Pattern ID	ICE and the associated reason	ICE strategy and data handling	Reference group for modeling the distribution of missing or unobservable data
N/A	Intake of prohibited medications Any reason	**Composite strategy** Treatment failure (i.e., no change from baseline.) Discard the collected data after the intake of prohibited medication(s) for a certain time frame based in the medication class (See Section 6.5.3 for details) and assign the value of zero change from baseline.	N/A
1	Temporary treatment interruption Due to COVID-19 /administrative or any other treatment unrelated reasons	**Hypothetical strategy** If participant complied with treatment. Data collected for a specific assessment during/after a temporary treatment interruption will be discarded for that assessment Week only.	Participants, from the same arm, who comply** with treatment regimen and have observed data at the time point that requires imputation. **(same as the first supplementary analysis)**
2	Temporary treatment interruption Due to AE	**Treatment policy strategy** Utilize all the collected data in the analysis.	Participants, from the same arm, who have available data before and during the treatment interruption. MI will be used to model change from a time point before treatment interruption to a time point with treatment interruption (Copy Increment from Reference imputation strategy). **(same as the first supplementary analysis)**
3	Permanent treatment discontinuation (Due to AE/lack of efficacy)	**Treatment policy strategy** If participant remained untreated/on background therapy after permanent discontinuation of the randomized treatment, use the treatment policy strategy and	Missing values that occur after the ICE will be imputed using a similar MI model as used in the main analysis of the primary endpoint, i.e., the MI regression model will include auxiliary variables

Pattern ID	ICE and the associated reason	ICE strategy and data handling	Reference group for modeling the distribution of missing or unobservable data
		utilize the collected data.	indicating (yes/no) whether the participant continued on randomized treatment at each visit (Guizzaro et al., 2021). **(same as the first supplementary analysis)**
		Composite strategy Otherwise, if participant initiated an alternative treatment, use the composite strategy assuming treatment failure (i.e., no change from baseline). Discard data after the ICE and assign the value of zero change from baseline	N/A
4	Permanent treatment discontinuation. Due to COVID-19 /administrative or any other treatment unrelated reasons	**Hypothetical strategy** If participant did not discontinue the randomized treatment due to the specified reason. Any data collected after permanent treatment discontinuation will be discarded.	All other participants from the same arm (i.e., participants without a similar ICE at the time point that requires imputation). **(same as the first supplementary analysis)**

** Complying with treatment regimen is defined as having no ICEs (i.e., no early discontinuation of randomized treatment, no treatment interruption in the respective week (definition in Table 1–2) and no intake of prohibited meds).

5.4 Secondary Endpoints Analysis

5.4.1 Key Secondary Endpoints

The key secondary efficacy endpoints in this study are:

- Mean change in severity of moderate to severe HF from baseline to Week 4 (assessed by HFDD) (all regions except regulatory submission in the US)
- Mean change in severity of moderate to severe HF from baseline to Week 12 (assessed by HFDD) (all regions except regulatory submission in the US)
- Mean change in frequency of moderate to severe HF from baseline to Week 1 (assessed by HFDD)
- Mean change in PROMIS SD SF 8b total score from baseline to Week 12
- Mean change in MENQOL total score from baseline to Week 12.

5.4.1.1 Definition of Endpoints

5.4.1.1.1 Hot Flash Daily Diary (HFDD) related endpoints

- Mean change in severity of moderate to severe HF from baseline to Week 4 (assessed by HFDD) (all regions except regulatory submission in the US)
- Mean change in severity of moderate to severe HF from baseline to Week 12 (assessed by HFDD) (all regions except regulatory submission in the US)
- Mean change in frequency of moderate to severe HF from baseline to Week 1 (assessed by HFDD) (all regions)

The calculation of the mean change in severity of moderate to severe HF from baseline to Week 4 and 12 will be done as described in Section 5.3.1. The calculation of the mean change in frequency of moderate to severe HF from baseline to Week 1 will be done analogously. The calculation of the frequency of moderate to severe HF for Week 1 will be based on Days 2-8 on treatment, where Day 1 corresponds to start of treatment. Week 2 will be based on days 8-14, Week 3 will be based on days 15-21 and so on. Detailed information on the HFDD is provided in Section 6.2.1.

5.4.1.1.2 Patient-reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b (PROMIS SD SF 8b) related endpoint

- Mean change in PROMIS SD SF 8b total score from baseline to Week 12

Participants responses to the 8 single items of the instrument are scored on a 1-5 numeric rating scale and aggregated to derive total raw scores ranging from 8-40. These total raw scores are then converted into T-scores using a look-up table for comparison with population norms. See further details in Section 6.2.2.

5.4.1.1.3 Menopause Specific Quality of Life Scale (MENQOL) related endpoint

- Mean change in MENQOL total score from baseline to Week 12

Participants responses to the 29 items will be scored individually on a scale indicating whether the participant experienced the symptom (yes/no). If they select yes, they rate how bothered they were by the symptom using a 6-point verbal descriptor scale. The items assess four domains of symptoms and functioning: vasomotor symptoms, psychosocial functioning, physical functioning, and sexual functioning. Domain scores are calculated as mean of single item converted scores. The MENQOL total score is the mean of the domain mean scores. See details in Section 6.2.3.

5.4.1.2 Main Analytical Approach

The same clinical questions are posed in respect of each primary and key secondary endpoint, and hence the data handling rules selected in respect of the ICEs are the same (detailed in Section 5.3.2). The key secondary endpoints will be analyzed analogous to the main analysis of primary endpoints (see Section 5.3.2). For PROMIS SD SF 8b T-scores are used. For MENQOL week 1 data will not be collected. The analysis will be adjusted accordingly.

PROMIS SD SF 8b raw scores for each item will be summarized by a frequency table with the number of observations and percentage by treatment group and by week (baseline, Week 1, 2, 3, 4, 8, 12, 16, 26 & 30). Bar charts will also be created by treatment group and over time.

PROMIS SD SF 8b total scores (converted as T-scores), total raw scores and MENQOL total scores will be summarized descriptively by treatment group and week (baseline, Week 4, 8, 12, 16, 26 & 30; for PROMIS in addition Week 1, 2, 3) using the number of non-missing observations, arithmetic mean and standard deviation, median, minimum and maximum. The absolute values at various weeks and the corresponding change from baseline per week will be summarized accordingly. The change from baseline over time will be shown using line plots for means together with 95% CIs by treatment group.

Graphical summaries with the absolute change from baseline on the x-axis and proportion of participants achieving this change on the y-axis, by treatment group, for week 4 and 12 will be presented for the MENQOL total score and for the PROMIS SD SF 8b total T-score.

5.4.1.3 Sensitivity Analyses

The sensitivity analyses as described in Section 5.3.3 for the primary endpoints will also be performed for the key secondary endpoints.

For tipping point analysis, the adjustments will be applied with the following delta values:

- PROMIS SD SF 8b total T-score, delta values of 3, 6, 9, etc.
- MENQOL total score, delta values of 0.5, 1, 1.5, etc.

5.4.1.4 Supplementary Analyses

The two supplementary estimands defined for the primary endpoints will also be used for the key secondary endpoints (see Section 5.3.4).

5.4.2 Supportive Secondary Endpoint(s)

The supportive secondary endpoints in this study are:

- Mean change in frequency of moderate to severe HF from baseline over time (assessed by HFDD)
- Mean change in BDI-II total score from baseline to Week 12
- Mean change in BDI-II total score from baseline to Week 26

The frequency of moderate to severe HF and the change from baseline will be summarized using descriptive statistics as described in Section 5.3.2.

The BDI-II consists of 21 items to assess the severity of depression over the past 2 weeks. Each item is scored from 0 to 3 and the total score ranging from 0 to 63 is calculated by summing up the ratings of the 21 items. Further details can be found in Section 6.2.4.

The BDI-II score will be summarized descriptively over time (i.e., at Week 4, 8, 12, 16, 26 & 30).

Further details with regards to the frequency of HF and baseline values can be found in Section 5.3.1.

5.5 Exploratory Endpoints Analysis

For all exploratory endpoints, the ICEs will be handled according to the treatment policy. No missing data imputation is planned for these endpoints.

5.5.1 HFDD related exploratory endpoints

The following exploratory endpoints based on the HFDD will be analyzed by descriptive statistical methods as described in Section 5.1 by treatment group:

- Proportion of participants with at least 50% reduction in frequency of (moderate to severe) HF at week 4
- Proportion of participants with at least 50% reduction in frequency of (moderate to severe) HF at week 12
- Time to specified treatment response
- Mean change in frequency of mild, moderate, and severe HF from baseline over time (assessed by HFDD)
- Mean change in frequency of nighttime awakenings from baseline over time (assessed by HFDD)
- Mean change in proportion of days with participants rating of "quite a bit" or "very much" sleep disturbances experienced due to HF from baseline over time (assessed by HFDD).

The proportion of participants with at least 50 % reduction in frequency of (moderate to severe) HF at week 4 and 12 will be calculated as the proportion of participants whose baseline value for the frequency of moderate to severe HF is reduced by ≥ 50 % at week 4 and 12, respectively.

A graphical summary with % change (0 to 100) in the frequency and severity of moderate to severe HF from baseline on the x-axis and proportion of participants achieving this change on the y-axis, by treatment group, for week 4 and 12 will also be provided. Similarly, a graphical summary with absolute change in the frequency and severity of moderate to severe HF from baseline on the x-axis and proportion of participants achieving this reduction on the y-axis, by treatment group, for week 4 and 12 will also be produced.

To assess the treatment response, a threshold value of 50 % reduction from the baseline of the frequency of moderate to severe HF will be considered. Time to first occurrence (i.e., the first week after baseline) of a reduction from the baseline value by 50 % will be analyzed. If the required treatment reduction by week 12 is not observed, the participant will be censored at week 12. Participants who drop out of the study before achieving the required reduction will be censored at the time of dropping out (i.e., at the last valuable week, defined as a week with at least 5 days of diary data, before dropping out). Kaplan-Meier estimates of the time to treatment response including a 95% confidence interval (calculated based on a normal approximation and using the Greenwood formula for the standard deviation) and cumulative incidence plots based on the Kaplan-Meier estimator will be presented by treatment group. Furthermore, the proportion of treatment responders will be summarized by treatment group. In addition, the proportion of treatment responders will be presented for the elinzanetant arm at Week 26.

The mean change in frequency of mild, moderate, and severe HF from baseline will be calculated similar to the mean change in frequency of moderate to severe HF described in Section 5.3.1, but will also include mild HF in the calculation. Specifically, the baseline value of the frequency of mild, moderate, and severe HF will be calculated by aggregating the available days during the 14 days prior to start of treatment to a mean daily frequency as (total number of mild, moderate, and severe HF during the 14 days prior to start of treatment) / (total number of available days with data). As per inclusion criteria 9 (see protocol), at least

11 days will be available for the derivation of the baseline value. Similarly, the frequency of mild, moderate, and severe HF for each week during the treatment period will be calculated using the available data during that particular week and aggregated to a mean daily frequency as (total number of mild, moderate, and severe HF during that week) / (total number of available days with data during that week). In case data is missing for more than 2 days within a week, the value for that week will be set to missing and no imputation of missing values will be performed. The mean change from baseline will then be calculated as the difference in the mean daily frequency of mild, moderate, and severe HF for a given week from the corresponding baseline value.

Summary statistics for the frequency of mild, moderate, and severe HFs as well as the absolute change from baseline will be presented by week and by treatment group. The change from baseline over time in the frequency of mild, moderate, and severe HFs will be shown using line plots for means together with 95% CIs by treatment group.

The mean change in frequency of nighttime awakenings from baseline over time is based on the number of nighttime awakenings during the previous night that is being assessed every morning as part of the HFDD. The baseline value of the nighttime awakenings will be calculated by averaging the nighttime awakenings from the available days during the 14 days prior to start of treatment to a mean daily frequency as (total number of nighttime awakenings during the 14 days prior to start of treatment) / (total number of available days with morning data). Similarly, the frequency of nighttime awakenings for each week during the treatment period will be calculated using the available data during that particular week and averaged to a mean daily frequency as (total number of nighttime awakenings during that week) / (total number of available days with morning data during that week). In case morning data is missing for more than 2 days within a week, the value for that week will be set to missing and no imputation of missing values will be performed. The mean change from baseline will then be calculated as the difference in the mean daily frequency of nighttime awakening for a given week from the corresponding baseline value. A graphical summary (i.e., cumulative distribution function) with absolute change (0 to 100) in the frequency of nighttime awakenings from baseline on the x-axis and proportion of participants achieving this reduction on the y-axis, by treatment group, for week 4 and 12 will also be provided.

Summary statistics for the frequency of nighttime awakenings as well as the absolute change from baseline will be presented by week and by treatment group. The change from baseline over time in the frequency of nighttime awakenings will be shown using line plots for means together with 95% CIs by treatment group.

Sleep disturbances due to HF are assessed every morning as part of the HFDD. Participants provide a rating of their sleep disturbances due to HF in the previous night as

- ‘not at all’
- ‘a little bit’
- ‘somewhat’
- ‘quite a bit’
- ‘very much’

To calculate the mean change in proportion of days with quite a bit or very much sleep disturbance due to HF, the number of days with participants having reported “quite a bit” or “very much” sleep disturbance will be used. The baseline value of the proportion of days with

quite a bit or worse sleep disturbances due to HF will be calculated based on the available days during the 14 days prior to start of treatment as (number of days with sleep disturbance due to HF rated as 'quite a bit' or 'very much')/ (total number of available days with morning data). The on-treatment value for the sleep disturbances due to HF for each week will be calculated using the available data during that particular week as (number of days with sleep disturbance due to HF rated as 'quite a bit' or 'very much' during that week) / (total number of available days with morning data during that week). In case morning data is missing for more than 2 days within a week, the value for that week will be set to missing and no imputation of missing values will be performed. Mean change in proportion of days with participants having reported "quite a bit" or "very much" sleep disturbance due to HF from baseline will then be calculated as the difference in the mean sleep disturbance value for a given week from the corresponding baseline value.

Summary statistics for the proportion of days with participants having reported "quite a bit" or "very much" sleep disturbance due to HF as well as the absolute change from baseline will be presented by week and by treatment group. The change from baseline over time of the proportion of days with participants having reported "quite a bit" or "very much" sleep disturbance due to HF will be shown using line plots for means together with 95% CIs by treatment group.

5.5.2 MenQoL Related exploratory endpoint

- Mean change in MENQOL domain and single item scores from baseline over time

Details on this instrument can be found in Section 5.4.1.1.3.

The absolute values and the change from baseline will be summarized by week (baseline and Week 4, 8, 12, 16, 26, and 30) and by treatment group using descriptive statistics as described in Section 5.1. The change from baseline over time of domain scores will be shown using line plots for means together with 95% CIs by treatment group.

5.5.3 Insomnia Severity Index (ISI)

- Absolute values and changes in the ISI total score over time.

The absolute values for the Total score and the change from baseline, as well as the severity ISI categories (i.e. "No clinically significant insomnia", "Subthreshold insomnia", "Clinical insomnia (moderate severity)" and "Clinical insomnia (severe)" as defined in Section 6.2.5) will be summarized for baseline and Week 4, 8, 12, 16, 26, and 30 by descriptive statistical methods as described in Section 5.1 by treatment group.

5.5.4 Patient Global Impression of Severity (PGI-S) and Change (PGI-C)

- Absolute values of the PGI-C individual item scores over time
- Absolute values and change in PGI-S individual item scores over time

The PGI-C includes 3 items using a 5-point response scale assessing the overall change in the frequency of HF ('much less' to 'much more'), the overall change in the severity of HF 'much better' to 'much worse') and the overall change in the severity of sleep problems ('much better' to 'much worse') since the participant started taking the study medication. Each item is answered independently.

The PGI-S includes 3 items using a 5-point response scale assessing the frequency of moderate to severe HF ('no HF' to 'very often'), the severity of HF ('no HF' to 'very severe'), and the severity of any sleep problems ('no sleep problem' to 'very severe') over the past 7 days. Each item is answered independently.

The answers to the 3 individual PGI-C items will be summarized in frequency tables for Week 4, 8, 12, 16, 26, and 30 by treatment group. The 3 individual PGI-S items will be summarized accordingly in frequency tables for baseline and Week 1, 2, 3, 4, 8, 12, 16, 26, and 30 by treatment group. The change from baseline in PGI-S individual scores will be analyzed using shift tables (observed frequencies at baseline versus post-baseline weeks).

5.5.5 European Quality of Life 5-dimension 5-level questionnaire (EQ-5D-5L)

- Absolute values and change in EQ-5D-5L single dimensions and health state visual analog scale (VAS) score over time

The EQ-5D-5L is a self-administered preference-based generic measure of health status which includes five dimensions:

1. mobility
2. self-care
3. usual activities
4. pain/discomfort
5. anxiety/depression.

Participants provide a rating for each question on a five-point Likert scale:

- for items 1-3:

  no problems, slight problems, moderate problems, severe problems, unable to do

- for item 4:

  no pain, slight pain, moderate pain, severe pain, extreme pain

- for item 5:

  not anxious or depressed, slightly anxious or depressed, moderately anxious or depressed, severely anxious or depressed, extremely anxious or depressed.

In addition, participants are asked to self-rate their own health today on a vertical 0 - 100 unit visual analogue scale (VAS), with 0 corresponding to "the worst health you can imagine", and 100 corresponding to "the best health you can imagine".

The answers to the 5 dimensions will be summarized in frequency tables for baseline and Week 1, 2, 3, 4, 8, 12, 16, 26, and 30 by treatment group. The change from baseline will be described using shift tables. The health state VAS values and the change from baseline will be analyzed for the same weeks by descriptive statistical methods as described in Section 5.1 by treatment group.

5.5.6 Beck Depression Inventory (BDI-II)

- Absolute values and changes in the BDI-II total score over time

Calculation of the BDI-II total score is described in Section 6.2.4. The absolute values and the change from baseline will be summarized for baseline and Week 4, 8, 12, 16, 26, and 30 using descriptive statistics as described in Section 5.1 by treatment group. The change from baseline in categories from BDI-II total score (0-13: none to minimal depression, 14 – 19 mild depression, 20 – 28 moderate depression and 29 – 63 severe) will be described using a shift table.

5.6 Safety Analyses

5.6.1 Extent of Exposure

The analyses described in this section will be presented for the SAF, and the FAS.

Treatment duration will be defined as the number of days from the day of first study drug intake up to and including the day of last study drug intake and will be summarized using descriptive statistics by treatment group, and by study drug. For the display by Week 1-12, the treatment duration ends on the end of treatment phase 1 (as defined in Section 5.1.2), or on the day of last study drug intake, whatever is earlier. For the display by Week 13-26, the treatment duration starts on the first day of treatment phase 2 (as defined in Section 5.1.2).

The extent of exposure to elinzanetant will be summarized as the total amount of study drug intake in grams and the average daily dose in mg using descriptive statistics per treatment group.

The compliance (as percentage) will be calculated as:

100 * Number of capsules taken / Number of planned capsules

The number of planned capsules is calculated as:

treatment duration * 2

All capsules, including the placebo capsules, will be counted. For participants who withdraw prematurely from the study drug, compliance will be calculated up to the time of last dose. The compliance will be summarized descriptively by treatment group. In addition, percentage of compliance will be categorized into 3 groups, less than 80%, 80 to 120% and greater than 120%, and the categories will be summarized by treatment group, and by study drug.

Treatment duration will be presented based on the data collected via the eCRF. If the date for end of exposure is not available in eCRF, the last day of drug intake from eDiary will be used to determine the end of exposure. Extent of exposure and compliance will be calculated and presented twice, once based on the ePRO daily instrument ‘Study drug intake documentation’ and once based on the eCRF ‘Drug Accountability’ and ‘Drug Exposure’ pages. The summaries for treatment duration, exposure and compliance will be presented for Week 1-12, Week 13-26 and overall, (i.e., the entire 26 weeks).

5.6.2 Adverse Events

Adverse events (AEs) will be collected from the start of study intervention until the last follow-up. (Serious) Adverse Events ((S)AEs) which are related to protocol required study procedures (e.g., (S)AE related to invasive study procedures) will be recorded as (S)AEs from the signing of the ICF (pre-treatment (S)AEs). Any medical occurrences/conditions that begin in the period between signing Informed consent form (ICF) and the start of study intervention, and which are not related to a protocol-required study procedure, will be recorded on the Medical History/Current Medical Conditions, not as AEs.

All AEs will be summarized using the Medical Dictionary for Regulatory Activities (MedDRA) preferred terms (PT) grouped by system organ class (SOC) (the current version at the time of analysis) and classified into pre-treatment AEs, treatment emergent (TEAE) and post-treatment AEs. Pre-treatment and post-treatment AEs will be shown by treatment group. Descriptive analysis for TEAEs will be performed for Week 1-12, Week 13-26, and overall, (i.e., the entire 26 weeks) by study drug.

A TEAE is defined as any adverse event occurring or worsening on or after the date of the first dose of study drug up to 14 days after the date of the last dose of study drug. AEs will be flagged as TEAE except for AEs for which there is clear evidence that the AE starts before date of first study drug intake (pre-treatment AEs) or after the date of last study drug intake + 14 days (post-treatment AEs).

If the severity of a previously reported event worsens or if the relationship to study intervention/protocol required procedures changes, two separate events should be reported.

Worsening of an AE is defined as follows:

- AE intensity is worsened (e.g., moderate to severe)
- AE changed to a serious event
- AE ends with death

In case an AE starts before the date of last study drug intake + 14 days and worsens after last study drug intake +14 days, it will be considered as two AEs, a TEAE and a post-treatment AE.

Partially missing onset AE date will be imputed following a worst-case approach as described in Section 5.1.1.

In case of events with different intensity within a participant, the maximum reported intensity will be used. If the same event is considered as both unrelated and related to the study drug within a participant, the event will be reported as related to study drug. If the drug relationship is missing, the event will be considered as being related to the study drug.

The tabulation will follow in principle the Bayer Global Standard Tables catalogue (v4.0 or later). Serious Adverse Events (SAEs) and Adverse events of special interest (AESI) will be summarized in the same way as described for TEAEs. A separate table summarizing TEAEs that occurred in more than 5% of the participants will be provided.

In addition, participants with a COVID-19 as an adverse event will be listed.

Adverse events of special interest (AESI)

AESIs will be identified by Standardized MedDRA Query (SMQ) searches, PTs and via eCRF as further detailed in Section 6.7. The following AEs are considered:

1. Any condition triggering close liver observation (as listed in Protocol Section 10.5)
2. Somnolence or fatigue
3. Phototoxicity
4. Post-menopausal uterine bleeding (women without hysterectomy are considered)

5.6.3 Additional Safety Assessments

5.6.3.1 Physical examination

Weight, hip and waist circumference and waist to hip ratio will be summarized by treatment group for each visit, including change from baseline where appropriate, using descriptive statistics as described in Section 5.1. The absolute values by visit and by treatment group will be shown using boxplots.

5.6.3.2 Vital signs

Vital signs (pulse rate, systolic blood pressure and diastolic blood pressure) will be summarized by treatment group for each visit, including change from baseline where appropriate, using descriptive statistics as described in Section 5.1. The absolute values by visit and by treatment group will be shown using boxplots. In addition, for systolic and diastolic blood pressure the line plots depicting the mean change from baseline ±95% CI by treatment group as well as scatter plot showing blood pressure values at end of treatment (EoT) against baseline by treatment group will be provided.

5.6.3.3 Electrocardiograms

Unscheduled electrocardiograms ECGs will be shown in data listings only. Bazett's formula, Fridericia's formula and/or calculations based on linear regression techniques will be shown for corrected QT-interval (QTc), when available.

5.6.3.4 Clinical laboratory

The tabulation of clinical laboratory data will be presented by treatment group for each visit according to the Bayer Global Standard Tables catalogue (version 4.0 or later). It includes:

- Summary statistics of continuous laboratory parameters, and their changes from baseline
- Frequency tables for categorical laboratory parameters

The distribution of laboratory values by treatment group for each visit will be presented graphically with boxplots. Line plots depicting the mean change from baseline ± 95% CI will be provided for CK, LDH, estradiol, HbA1c, total cholesterol, HDL, LDL and TG and for liver parameters: AST, ALT, ALP, Total bilirubin and INR.

Treatment-emergent (i.e., 14 days from the last drug intake) high and low laboratory abnormalities will be summarized by treatment group. Both scheduled and unscheduled laboratory measurements will be used in treatment-emergent abnormal laboratory tables.

Central laboratory will be considered for descriptive analysis. Local laboratory measurements will be listed, if available.

5.6.3.4.1 Liver monitoring

The following parameters will be investigated in addition to the standard lab presentations:

- Aspartate aminotransferase (AST) (in U/L),
- Alanine aminotransferase (ALT) (in U/L),
- Alkaline phosphatase (ALP) (in U/L),
- Total bilirubin in serum (TB) (in mg/dL)

- International normalized ratio (INR).

Frequency tables, presenting number and percentage of participants by treatment group for baseline and post-baseline values (both scheduled and unscheduled measurements can be used), will be presented for the following categorizations (where ULN stands for "Upper Limit of Normal"):

- For ALT and AST, separately:
  - ≥1×ULN, ≥3×ULN, ≥ 5×ULN, ≥ 8×ULN, ≥ 10×ULN, ≥ 20×ULN
- For ALT and AST combined (if at least one of ALT and AST falls into the category):
  - ≥ 3×ULN, ≥ 5×ULN, ≥8×ULN, ≥ 10×ULN, ≥ 20×ULN
- For Total bilirubin:
  - ≥1×ULN, ≥2×ULN, ≥ 5×ULN, ≥8×ULN
- For ALP:
  - ≥1.5×ULN, ≥2.0×ULN, ≥3.0×ULN
- For INR:
  - ≥1.5, ≥2

Frequency tables presenting number and percentage of participants by treatment group for post-baseline will be presented for the following combinations of ALT/AST and Total bilirubin or INR,

Relative to ULN:

- ALT or AST ≥3×ULN and ≥1.5×ULN in Total bilirubin
- ALT or AST ≥ 3×ULN and ≥2×ULN in Total bilirubin
- ALT or AST ≥3×ULN followed by ≥2×ULN in Total bilirubin (measured within 30 days afterwards) (Hy's Law criteria).
- ALT or AST ≥3×ULN and ≥1.5 of INR
- ALT or AST ≥5×ULN for more than 2 weeks
- ALT or AST ≥8×ULN
- ALT or AST ≥ 3×ULN with the appearance of any signs or symptoms based on close liver observation eCRF page.
- ALP ≥ 2×ULN and Total Bilirubin ≥2×ULN

Relative to baseline (BL):

- ALT or AST ≥3×BL and Total bilirubin ≥2×BL
- ALT ≥3×BL or AST ≥3×BL
- ALT ≥5×BL or AST ≥5×BL
- ALT or AST ≥2×BL and Total bilirubin ≥2×BL

- ALP ≥2×BL and Total bilirubin ≥2×BL

Time to event analysis

Cumulative incidence estimates for the time to first occurrence of ALT ≥ 3×ULN and first occurrence of ALP ≥ 3×ULN will be derived. If no such an increase is observed, the observation is censored at the last visit date. Tables with the number of participants under risk, cumulative number of participants with ALT ≥3×ULN and ALP ≥ 3×ULN, and estimated probability for an event including 95% CIs (two-sided) will be presented. Furthermore, cumulative incidence curves will be provided by treatment group. If a participant does not have any post-baseline data, she will be censored at baseline.

Figures

Hepatocellular drug-induced liver injury (DILI) screening plot will be displayed to identify cases of possible serious hepatocellular DILI. In the plot each patient is plotted based on their maximum postbaseline Total bilirubin (y-axis) and transaminase values (ALT or AST, whichever is higher. Together with the plot a frequency table for patients in each quadrant (Potential Hy's Law, Cholestasis, Temple's corollary) will be shown.

Cholestatic drug-induced liver injury screening plot will be shown to identify significant ALP elevation in the setting of hepatic dysfunction. Maximum postbaseline Total bilirubin is plotted against maximum postbaseline ALP. Similar to above a frequency table for patients in each quadrant (Total bilirubin ≥ 2xULN and ALP ≥ 2xULN, Total bilirubin ≥ 2xULN and ALP < 2xULN, Total bilirubin < 2xULN and ALP ≥ 2xULN) will be shown.

Individual patient presentations

If a patient has ALT ≥ 3×ULN at any time point, a plot for her individual time course in the following laboratory parameters will be presented: 1) ALT, AST, Total bilirubin, and ALP relative to ULN over time and 2) absolute values for INR over time. It will be indicated within the plot on which days the study drug was taken (i.e. start and stop dates of treatment intake). Furthermore, listings will be provided for INR and with results relative to ULN for liver-related parameters, i.e. ALT, AST, Total bilirubin, ALP.

5.6.3.5 Pregnancy test

Pregnancy test results will be listed.

5.6.3.6 Mammogram

A frequency table for mammogram findings (normal/abnormal) at screening (as baseline value) and at EoT will be produced by treatment group.

5.6.3.7 Transvaginal ultrasound

Number and percentage of participants with ultrasound performed and result of overall assessment at each visit (baseline and EoT) will be presented by treatment group.

Endometrial thickness will be summarized by treatment group for each visit, including change from baseline, using descriptive statistics for women without hysterectomy.

Number and percentage of participants that developed ovarian cyst during the study will be presented by treatment group.

5.6.3.8 Cervical cytology

A frequency table for cervical smears findings at screening will be provided by treatment group.

5.6.3.9 Endometrial biopsy

Analyses will present the number and percentage of either participants or biopsies for women without hysterectomy. Both scheduled and unscheduled biopsy assessments will be considered.

Presentations will be done for

- the Majority read
- All reads (Reader #1 - #3)

Majority read: Majority read will be determined for main results and subcategories (see Table 5–8). First, adequacy (for part II and III) of tissue will be investigated by all readers. If at least 2 of the 3 readers consider the tissue adequate/sufficient, the majority for the main results will be assessed. If there is a majority with respect to the main result, majority of the respective subcategories will be determined. If no majority result is available (3 different results in 3 readers, 2 different results in 2 readers), either “no consensus” or the worst case will be presented. Table 5–8 presents an overview of the biopsy results including the approach which is used in case no majority is available. Other observations will be listed only.

All read: All read results will be based on all biopsies with results from at least one reader.

Table 5–8: Overview of biopsy endpoints including majority result handling and worst case

Part		Endpoint	No majority available
Main results			
I		Adequate endometrial tissue	- *(not possible)*
II	Main diagnosis	Benign endometrium Endometrial Hyperplasia (WHO 2014 classification) Malignant Neoplasm	Worst case: List is ordered by severity, from low to high
III		Endometrial Polyp	Worst case: yes
Subcategories			
II	Main diagnosis	Benign endometrium (select one) •Atrophic • Inactive • Proliferative • Disordered Proliferative • Secretory • Menstrual • Endometritis • Other, specify	"no consensus"
		Endometrial Hyperplasia (WHO 2014 classification) (select one) • Hyperplasia without atypia • Atypical hyperplasia / Endometrioid Intraepithelial Neoplasia (EIN)	Worst case: Atypical hyperplasia / Endometrioid Intraepithelial Neoplasia
		Malignant Neoplasm • Endometrial Neoplasm • Other Malignant Neoplasm	"no consensus"
III		Endometrial Polyp (select one) • Atrophic • Functional • Hyperplastic	"no consensus"

5.6.3.10 Sleepiness scale

The sleepiness scale is a sponsor developed questionnaire containing 3 items assessing the degree of sleepiness experienced by the participant in the morning, in the afternoon and in the evening of the same day, using the electronic handheld device together with the HFDD evening diary assessment. In case one entry on the sleepiness scale was missed, retrospective data entry is possible for a pre-defined period of time (the entry option will be available between midnight until 10:59AM on the day after the missed entry. During the screening period assessments will be done daily. During the treatment period the assessment will be done on 7 consecutive evenings during Weeks 1, 4 and 12.

Participants respond to the items using a 5-point verbal rating scale (from "0" = not at all, to "4" = very much). Daily sleepiness score will be calculated by averaging daily individual item scores (i.e., morning, afternoon, and evening scores).

Descriptive summaries for 7-day average of daily individual single item scores and 7-day averages of a daily sleepiness score and corresponding change from baseline values will be created by treatment group for each week. 7-day averages can be derived, if daily score at least 5 out of 7 days is available, otherwise average score will be set as missing.

The baseline sleepiness score will be derived from days with available data within the 14-day period before randomization. A minimum of 11 days should be available for the derivation of the baseline value.

5.6.3.11 Electronic Columbia-Suicide Severity Rating Scale (eC-SSRS)

Suicidal ideation and behavior will be monitored by eC-SSRS questionnaire. The eC-SSRS assesses lifetime suicidality during an initial baseline evaluation, and then prospectively monitors ideations and behaviors at subsequent follow-up assessments. The eC-SSRS assesses the severity and frequency of suicidal ideation and behavior.

eC-SSRS outcomes will be summarized by treatment group for each visit (baseline, Week 4, 8, 12, 16, 26, and 30) using descriptive statistics for continuous variables and frequency tables for categorical outcomes as described in Section 5.1.

5.7 Other Analyses

Other pre-specified objectives in this study are:

- To evaluate variability in exposure in relation to the efficacy and safety for elinzanetant
- To further investigate elinzanetant (e.g., mode-of-action-related effects, safety) and to further investigate pathomechanisms deemed relevant to VMS and associated health problems

These will be evaluated accordingly by:

- Systemic exposure of elinzanetant in plasma via sparse PK sampling
- Various biomarkers (e.g., diagnostic, safety, pharmacodynamic, monitoring, or potentially predictive biomarkers)

5.7.1 Other Variables and/or Parameters

5.7.1.1 Pharmacokinetics

PK analyses will be described in a separate document and the results will be presented outside CSR. Only concentration data will be provided in a listing (Section 10).

5.7.1.2 Genetics

Genetic as well as non-genetic analyses may be part of the biomarker investigations in this study, if approved by local Ethics committees / Institutional Review Boards and competent authorities.

Pharmacogenetic investigations may be of any kind, except for whole genome sequencing.

The analysis will be described in a separate document and the results will be presented outside CSR.

5.7.1.3 PD /safety biomarkers from blood

The analyses for PD/safety biomarkers from blood will be described in a separate document and the results will be presented outside CSR.

5.7.1.4 Psychometric properties of selected questionnaires

Data from the study may be used to assess psychometric properties of scores from HFDD, PROMIS SD SF 8b, MENQOL and sleepiness scale to be described in a separate document and the results will be presented outside CSR.

5.7.2 Subgroup Analyses

Exploratory subgroup analyses using descriptive statistics will be provided for the primary and key secondary endpoints for the following subgroups:

- Region (North America vs. rest of the world)
- Race
- Ethnicity
- BMI (<18.5, 18.5 to <25, 25 to <30, >= 30 kg/m^2)
- Smoking history (Never, Former, Current; derived from habitual cigarette smoking and any other tobacco/nicotine from the CRF)

For the key secondary endpoints based on the PROMIS SD SF 8b the above exploratory subgroup analyses will only be conducted for the total T-scores. For the MENQOL, they will be presented for the total scores and the vasomotor symptoms subdomain score.

Descriptive statistics for PROMIS SD SF 8b total T-scores will be provided for the following subgroups based on Insomnia Severity Index (ISI) administered at baseline (see Section 5.5.3 more details on ISI):

- 0-14 = No clinically significant and subthreshold insomnia
- 15–21 = Clinical insomnia (moderate severity)
- 22–28 = Clinical insomnia (severe)

5.8 Interim Analyses

No interim analysis is planned for this study.

6. Supporting Documentation

6.1 Appendix 1: List of Abbreviations

Abbreviation	Definition
AE	Adverse event
AESI	Adverse events of special interest
AIC	Akaike's information criteria
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AR(1)	Autoregressive first order
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical classification
BDI	Beck Depression Inventory
BL	Baseline
BMI	Body mass index
BRM	Blind review meeting
CFB	Change from baseline
CI	Confidence interval
CM	Concomitant medications
CMH	Cochran-Mantel-Haenszel
COVID-19	Coronavirus disease of 2019
CS	Compound symmetry
CSR	Clinical study report
DILI	Drug-induced liver injury
ECG	Electrocardiogram
eC-SSRS	Electronic Columbia-Suicide Severity Rating Scale
eCRF(s)	Electronic case report form(s)
eDiary	Electronic diary
EMA	European Medicines Agency
ePROs	Electronic Participant-Reported Outcomes
EQ-5D-5L	European Quality of Life 5-dimension 5-level questionnaire
EoT	End of treatment
FAS	Full analysis set
FDA	Food and Drug Administration
GnRH	gonadotropin-releasing hormone
HF	Hot flash or Hot flashes
HFDD	Hot flash daily diary
ICE(s)	Intercurrent event(s)
ICF	Informed consent form
ICH	International Council on Harmonization
INR	International normalized ratio
ISI	Insomnia severity index
IUD	Intrauterine Device
LPLV	Last patient last visit
MAO-A	Monoamine oxidase A
MAR	Missing at random
MCMC	Monte Carlo Markov Chain
MedDRA	Medical dictionary for regulatory activities
MENQOL	Menopause-specific quality of life questionnaire
MI	multiple imputation
MMRM	Mixed model repeated measures
PD	Pharmacodynamic
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PK	Pharmacokinetics

PROMIS SD SF 8b	Patient-Reported Outcomes Measurement Information System Sleep Disturbance short-form 8b
PT	Preferred term
QTc	Corrected QT-interval
ROW	Rest of the World
SAEs	Serious Adverse Events
SAF	Safety analysis set
SAP	Statistical analysis plan
SD	Standard deviation
SE	Standard error
SMQ	Standardised MedDRA Query
SOC	System organ class
SSRI	Selective serotonin reuptake inhibitors
TB	Total bilirubin
TEAE	Treatment emergent adverse event
ULN	Upper Limit of Normal
UN	Unstructured
US	United States
VAS	Visual analog scale
VMS	Vasomotor symptoms

6.2 Appendix 2: Additional information on scoring and measurement properties of the HFDD, the PROMIS SD SF 8b, MENQOL, BDI-II, and ISI

6.2.1 HFDD

Participants' assessments of HF will be recorded electronically twice daily using the sponsor developed HFDD. The HFDD will be completed in the morning after waking up (morning diary) and each evening at bedtime (evening diary) on the hand-held device. The HFDD items assess the number of mild, moderate, and severe HF experienced during the day and during the night. In addition, the number of awakenings during the night and disturbance of sleep due to HF will be documented in the morning diary.

Mild HF are defined as a "sensation of heat without sweating", moderate HF are defined as a "sensation of heat with sweating, but able to continue activity", and severe HF are defined as a "sensation of heat with sweating, causing cessation (stopping) of activity".

6.2.2 PROMIS SD SF 8b

The PROMIS SD SF 8b includes 8 items assessing sleep disturbance over the past 7 days (Yu et al. 2011). Items assess sleep quality, sleep depth and restoration associated with sleep, perceived difficulties with getting to sleep or staying asleep and perceptions of the adequacy of and satisfaction with sleep. Participants respond to the items on a 5-point scale (scored from 1 to 5) from "not at all", "never", "very poor" to "very much", "always" or "very good". Four of the items are scored reversely. The individual item scores for a participant will be summed to derive a total raw score. Total raw scores range from 8 to 40, with higher scores indicating greater severity of sleep disturbance. Total raw scores will be converted into T-scores for comparison with population norms (United States general population). The conversion table is provided in Table 6–1. (PROMIS Sleep Disturbance – Scoring Manual). T-scores will be used in the confirmatory analysis.

In this study the PROMIS SD SF 8b will be applied electronically and responded to by the participants at home and during selected in person visits using the handheld device once weekly during the weeks 1-4 and thereafter at weeks 8, 12, 16, 26 and 30.

Table 6–1: Sleep Disturbance 8b - Conversion table

Sleep Disturbance 8b *Short Form Conversion Table*		
Raw Score	**T-score**	**SE***
8	28.9	4.8
9	33.1	3.7
10	35.9	3.3
11	38.0	3.0
12	39.8	2.9
13	41.4	2.8
14	42.9	2.7
15	44.2	2.7
16	45.5	2.6
17	46.7	2.6
18	47.9	2.6
19	49.0	2.6
20	50.1	2.5
21	51.2	2.5
22	52.2	2.5
23	53.3	2.5
24	54.3	2.5
25	55.3	2.5
26	56.3	2.5
27	57.3	2.5
28	58.3	2.5
29	59.4	2.5
30	60.4	2.5
31	61.5	2.5
32	62.6	2.5
33	63.7	2.6
34	64.9	2.6
35	66.1	2.7
36	67.5	2.8
37	69.0	3.0
38	70.8	3.2
39	73.0	3.5
40	76.5	4.4

*SE = Standard Error on T-score metric

6.2.3 MENQOL

The MENQOL questionnaire is comprised of 29 items assessing the presence of menopausal symptoms and the impact of menopause on health-related quality of life over the past week (Hilditch et al., 2008). For each item, the participant indicates if they have experienced the symptom (yes/no). If they select “yes”, they rate how bothered they were by the symptom using a six-point verbal descriptor scale, with response options ranging from 0 'not at all bothered' to 6 'extremely bothered'. The conversion table is provided in Table 6–2.

Table 6–2: MENQOL - Conversion table

Subject Response	Converted Score
The participant responded 'NO', she did not experience the problem	1
The participant experienced the problem and rated it as '0' on the bothered scale;	2
The participant experienced the problem and rated it as '1' on the bothered scale;	3
The participant experienced the problem and rated it as '2' on the bothered scale;	4
The participant experienced the problem and rated it as '3' on the bothered scale;	5
The participant experienced the problem and rated it as '4' on the bothered scale;	6
The participant experienced the problem and rated it as '5' on the bothered scale;	7
The participant experienced the problem and rated it as '6' on the bothered scale;	8

Domain scores are calculated by averaging the converted individual item scores (range 1-8) related to the respective domain. (Domains: Vasomotor - items 1 to 3, Psychosocial – items 4 to 10, Physical- items 11- to 26, Sexual – items 27 to 29.) For a MENQOL total score the aggregated mean of the mean domain scores is calculated. Higher scores indicate greater bother. MENQOL total score, domain scores and individual item scores will be summarized in this study.

In this study the MENQOL will be applied electronically and responded to by the participants at home and during selected in person visits at weeks 1, 4, 8, 12, 16, 26 and 30.

6.2.4 BDI-II

The BDI-II consists of 21 items to assess the severity of depression over the past 2 weeks. The 21 items are with respect to various symptoms of depression, specifically:

- Sadness
- Pessimism
- Past Failure
- Loss of Pleasure
- Guilty Feelings
- Punishment Feelings
- Self-Dislike
- Self-Criticalness
- Suicidal Thoughts or Wishes
- Crying
- Agitation
- Loss of Interest
- Indecisiveness

- Worthlessness
- Loss of Energy
- Changes in Sleeping Pattern
- Irritability
- Changes in Appetite
- Concentration Difficulty
- Tiredness or Fatigue
- Loss of Interest in Sex

Each item consists of a list of four response options arranged in increasing levels of severity, except for "Changes in Sleeping Pattern" and "Changes in Appetite" which contain 7 response options each. For the items with 4 response options, each item is rated on a 4-point verbal response scale ranging from 0 (not at all) to 3 (extreme form of each symptom). The items with 7 response options are also scored from 0 to 3 as

Response option 1 → score=0

Response option 2 or 3 → score=1

Response option 4 or 5 → score=2

Response option 6 or 7 → score=3.

The total score ranging from 0 to 63 is calculated by summing up the ratings to the 21 items.

6.2.5 ISI

The ISI is a seven-item instrument that quantifies the participant perception of insomnia severity, along with the impact of insomnia on daytime functioning in adults in the last two weeks (Bastien et al. 2001). The seven questions are with respect to:

- difficulty falling asleep
- difficulty staying asleep
- problems waking up too early
- satisfaction/dissatisfaction with current sleep pattern
- noticeability of sleep problems by others
- worries/distress caused by the sleep difficulties
- extent of interference of sleep difficulties with daily functioning.

It is scored on a five-point Likert scale from 0 to 4 depending on the item:

- Items 1-3: 0='none' to 4='very severe'
- Item 4: 0='very satisfied' to 4='very dissatisfied'
- Item 5: 0='not at all noticeable' to 4='very much noticeable'
- Item 6: 0='not at all worried' to 4='very much worried'
- Item 7: 0='not at all interfering' to 4='very much interfering' (Item 7)

The scores for each item are summed to produce the total score (maximum 28) which allow categorization of severity of insomnia.

Total score categories:

- 0–7 = No clinically significant insomnia
- 8–14 = Subthreshold insomnia
- 15–21 = Clinical insomnia (moderate severity)
- 22–28 = Clinical insomnia (severe)

6.3 Appendix 3: Pattern-mixture modeling approach for supplementary analysis

A pattern mixture modeling approach (Little 1993) will be used in combination with multiple imputation (Ratitch , O'Kelly et al., 2013; Guizzaro et al., 2021) to handle missing and unobservable data that occur in presence of ICEs in a way that aligns with the estimand strategy for each type of ICE.

We will use a reference-based pattern imputation approach which makes use of observed data from the reference group for estimating the multiple imputation model for each pattern that requires distinct assumptions about a plausible distribution of missing/unobservable data. The reference group consists of participants who are expected to have similar outcomes to those with missing/unobservable values after conditioning on baseline covariates, partially observed data, and, in some cases, indicators (yes/no) for whether participants remained on the randomized treatment at each visit. Pattern definitions and the corresponding reference groups that will be used for the supplementary analysis of the primary endpoints are specified in Table 5–5 and Table 5–7. Multiple imputation models will be estimated separately for frequency and severity of hot flashes (referred to by the generic term “outcome” below).

The pattern-mixture multiple imputation and analysis method will be implemented by a series of steps using SAS PROC MI, MIXED, and MIANALYZE:

1. Imputation: Impute missing and unobservable values at baseline, Week 1, 4, 8 and 12 (see Table 5–5 and Table 5–7 for definition of patterns):

Impute intermittent (non-monotone) missing values that occur outside of the time frame of any ICE as well as missing and unobservable values that occur in presence of ICEs in pattern 1.

A joint multivariate normal multiple imputation model will include outcomes at baseline, Week 1, 4, 8 and 12, as well as the fixed, categorical effects of treatment and region.

This step will be implemented using PROC MI with the MCMC statement and the following options: “CHAIN = MULTIPLE”, and “SEED = 21652” to produce 500 imputed datasets. The input dataset will contain all participants with missing/unobservable data targeted by this step as described above and participants from the reference group, i.e., participants who complied with the randomized treatment (please see the definition of compliance in Table 1–2). Partially imputed data will serve as input to complete imputation in patterns 3, and 4.

Impute missing values occurring during an ICE associated with pattern 2.

A multiple imputation regression model will be used to model change from a time point before treatment interruption to a time point during treatment interruption as the dependent variable (change equal to pre-interruption value minus value at the time point of interruption). The model will include as predictors the baseline value, value before treatment interruption, and the fixed, categorical effects of treatment and region. Because the number of participants in this pattern is expected to be small, modeling of change in outcome from pre-interruption to during interruption will be done by pooling data across time points where treatment interruptions occur. Missing values at a time point *k* will be replaced by subtracting the change value predicted by the imputation model from the participant's value at the time point *(k-1)* prior to treatment interruption.

This step will be implemented with PROC MI and MONOTONE REG statement explicitly specifying the model described above. The input dataset will include all participants in pattern 2 and participants with temporary interruption due to treatment unrelated reasons from pattern1 who have observed data before and during the interruption.

Impute missing and unobservable data occurring after an ICE associated with pattern 3.

Missing values that occur after the ICE will be imputed using a similar MI model as used in the main analysis of the primary endpoint. That is, the MI monotone regression model (PROC MI with MONOTONE REG statement) will include auxiliary variables indicating (yes/no) whether the participant continued on randomized treatment at each visit (Guizzaro et al., 2021) in addition to baseline value, values at previous time points, and the fixed, categorical effects of treatment and region. In case there are less than 10 participants (who permanently discontinued the randomized treatment) across both treatment groups with available post-discontinuation data at any intermediate time point, the indicator variable will not be included in the imputation model for that time point. The input dataset will include all participants in the analysis set, including their observed values post premature treatment discontinuation and values that might have been imputed as part of pattern 1 and 2. Prior to this step, intermittent missing values in the input dataset will be partially imputed using PROC MI with the MCMC statement using multiple chains.

Impute missing and unobservable data occurring during an ICE associated with pattern 4.

Unobservable data in this pattern will be monotone missing. They will be imputed using multiple imputation regression models for outcome at each affected time point. The model will include as predictors the baseline value, values at time points before treatment discontinuation, and the fixed, categorical effects of treatment and region. The input dataset will include all participants in the analyses set, their observed values and values imputed on previous steps outlined above. That is, the MI model for pattern 4 will reflect a mixture of participants, some of which fully comply with the randomized treatment and some with other types of ICEs.

If an unobserved/missing value is imputed as a value below 0, a post processing step will be applied to set the imputed value to 0.

In order to impute a missing value at baseline (as part of the imputation of intermittent missing values), a minimum of 2 post-baseline values (out of 4) should have been observed for the participant.
At the end of the four imputation steps, 500 datasets will be produced and each imputed dataset will be analyzed as described in the analysis step below.

2. Analysis: Estimate of the treatment difference based on each imputed dataset, using the analysis model for the primary endpoints (i.e. MMRM as specified in Section 5.3.2).

The estimated treatment difference at Week 1, 4, 8 & 12 from all imputed datasets will be saved and combined in the next step as described below.

3. Combining results of analyses from multiple imputed datasets: The results of the MMRM analysis on 500 imputed datasets will be combined using Rubin's rule to derive the final estimate for the treatment difference, its 95% CI and p-value (one-sided). This will be done using SAS PROC MIANALYZE.

6.4 Appendix 4: Tipping point analysis

A tipping point analysis will be applied by applying an unfavorable additive shift (referred to as delta adjustment) to values imputed to fill in the missing and unobservable data that occur in presence of ICEs. Delta adjustment will be applied only to imputed values in the elinzanetant arm. No delta adjustment will be applied to missing values that occur outside of the time frame of any ICE. The following steps will be implemented for each type of the primary endpoint, i.e., frequency and severity of HF:

1. The imputation step will be done as described in Section 5.3.2.
2. Prior to the analysis step, each imputed value in the elinzanetant arm that occurs in the time frame of an ICE would have a value of "delta" added to it (see Section 5.3.3 for settings of delta for each type of endpoint). The same value of delta will be applied at each time point.
3. Observed, imputed, and adjusted data will be analyzed and results combined as described in Section 5.3.2.
4. The steps for imputation, modeling, combining the results will be repeated with increasing settings of "delta" (see Section 5.3.3) for each primary endpoint until the estimated treatment difference at Week 4 or Week 12 is no longer statistically significant at the 0.025 one-sided level. The corresponding "delta" setting will be selected as the tipping point for that endpoint.

6.5 Appendix 5: Details regarding the identification of intercurrent events

6.5.1 Temporary treatment interruption

Per definition (see Section 1.1), the ICE "temporary treatment interruption" will only be considered for week 1, week 3 and 4, week 7 and 8, and week 11 and 12. Specifically, temporary treatment interruption is defined as

- Treatment taken on <5/7 days during week 1 for Week 1
- Treatment taken <80% during weeks 1-4 OR treatment taken on <5/7 days during either week 3 or 4 for Week 4

- Treatment taken <80% during weeks 1-8 OR treatment taken on <5/7 days during either week 7 or 8 for Week 8
- Treatment taken <80% during weeks 1-12 OR treatment taken on <5/7 days during either week 11 or 12 for Week 12.

The criterion for treatment taken on 5/7 days during the corresponding two weeks, as defined above, is met, if in any selected 7 consecutive days within those weeks, the treatment has been taken for less than 5 days (i.e., at least 3 missed intakes).

If the criterion is met, an ICE flag will be placed for the corresponding visit week for HFDD. For PROMIS and MenQoL, the ICE flag will be placed only if the questionnaire is filled in after the third drug missed intake day.

To identify whether < 80 % treatment was taken, planned compliance during weeks 1-4 will be calculated as

100 * Number of capsules taken / 56

where 56 is based on two tablets per day over 28 days. Planned compliance during weeks 1-8 will be calculated as

100 * Number of capsules taken / 112

where 112 is based on two capsules per day over 56 days.

Planned compliance during weeks 1-12 will be calculated as

100 * Number of capsules taken / 168

where 168 is based on two capsules per day over 84 days.

To calculate the number of days per week with treatment intake, a day where at least one capsule was taken, will be considered as a day with treatment intake.

For the identification of the ICE temporary treatment interruption, the compliance and daily treatment intake will be calculated based on the ePRO instrument ‘Study drug intake documentation’.

Regarding the reasons, AEs that lead to a treatment interruption, together with start and end date, will be collected on the respective eCRF page. Similarly, reasons related to COVID-19 or administrative reasons will also be collected in the eCRF.

The reasons for temporary treatment interruption will be reviewed during blind review of the data. The observed temporary treatment interruptions together with the corresponding reason will be presented in data listings.

6.5.2 Permanent discontinuation of randomized treatment

The reasons for permanent discontinuation of randomized treatment will be collected on the corresponding eCRF page, among others with AE or lack of efficacy as possible options. A permanent discontinuation of randomized treatment will be considered to be due to “other treatment unrelated reasons” if AE or lack of efficacy was not selected as reason and any of the following was selected:

- Non-compliance with study device
- Site terminated by sponsor

- Technical problems
- Pregnancy
- Subject decision
- Subject decision: unwillingness to comply with study procedures
- Lost to follow-up
- Other
- COVID-19 pandemic related: subject specific
- COVID-19 pandemic related: other
- Non-compliance with study drug
- Randomized by mistake
- Withdrawal by subject

The observed permanent discontinuations of randomized treatment together with corresponding reason will be presented in data listings.

Discontinuation due to AE or lack of efficacy may be handled differently depending on estimand for participants who remained untreated/background therapy and for participants who initiate alternative VMS treatment. The treatment status (i.e., remained untreated/on background therapy vs. initiate alternative VMS treatment) should be captured for all participants who discontinued from randomized drug irrespective of the reason for discontinuation.

Table 6–3: Preliminary list of alternative VMS treatment by drug grouping (can be updated during blind review meeting)

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2	Duration of impact with respect to efficacy
		5	Hormone replacement therapy	2	Oestrogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous or intradermal.

		5	Hormone replacement therapy	3	Progestogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous or intradermal.
1633	Drugs for psychiatric disorders	111	Antidepressants	113	Selective serotonin reuptake inhibitors (SSRI)	From 1st dose to 4 weeks after the last dose

Table 6–4: Preliminary list of alternative VMS treatment by drug names (can be updated during blind review meeting)

WHO Drug Name	WHO Drug Record Number	WHO Drug Sequence Number 1	WHO Drug Sequence Number 2	Duration of impact with respect to efficacy	
OXYBUTIN	005389	02	057	From 1st dose to 4 weeks after the last dose	

Although more alternative treatment options may be available, only drug groupings/ drug names are listed that were reported in OASIS 1-3 studies as of finalization of this statistical analysis plan.

6.5.3 Intake of prohibited concomitant medication having impact on efficacy

Drug groupings, individual drugs and manual review will be used to identify prohibited concomitant medications that were defined in the protocol to influence efficacy. All here listed drug groupings/ drug names were reported in OASIS 1-3 studies as of finalization of this statistical analysis plan, although other medications may also be prohibited per the protocol.

In terms of the efficacy endpoints related to the change from baseline in HF frequency and HF severity:

1. For the change from baseline at Week 1 any intake of prohibited medication during Week 1 that has an effect on HF will be considered an intercurrent event.
2. For weeks 4, 8, and 12, not only intake in the respective week but also, intake that occurred prior to week 4, 8, and 12 should be considered as an intercurrent event if within the washout period of the prohibited drug.

A similar approach will be used for the endpoints related to sleep quality (mean change in PROMIS SD SF 8b total score from baseline) and to the menopause related quality of life (mean change in MENQOL total score from baseline) with respect to any prohibited medication that has an effect on sleep quality and menopause related quality of life, respectively.

The list of prohibited concomitant medication together with the pre-defined washout time period for their effect will be reviewed during blind review of the data and finalized prior to unblinding.

Table 6–5: Preliminary list of prohibited concomitant medication by drug grouping (can be updated during the blind review meeting)

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2	Duration of impact with respect to efficacy (considered up to 12 weeks*)
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	109	GnRH agonists	From 1st dose to 12 weeks after the last dose
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	110	GnRH antagonists	From 1st dose to 12 weeks after the last dose
		5	Hormone replacement therapy	2	Oestrogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous or intradermal.
		5	Hormone replacement therapy	3	Progestogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous or intradermal.
		772	Cancer therapies	738	Endocrine antineoplastic therapy	From 1st dose to 12 weeks after the last dose

1633	Drugs for psychiatric disorders	111	Antidepressants	114	Monoamine oxidase (MAO) inhibitors, non-selective	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	115	Monoamine oxidase A (MAO-A) inhibitors	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	112	Non-selective monoamine reuptake inhibitors	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	113	Selective serotonin reuptake inhibitors (SSRI)	From 1st dose to 4 weeks after the last dose
45	Drugs interacting with CYP3A	240	CYP3A inducers	225	Moderate CYP3A inducers	From 1st dose to 4 weeks after the last dose
45	Drugs interacting with CYP3A	240	CYP3A inducers	265	Strong CYP3A inducers	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	1830	Antidepressant Serotonin Norepinephrine Reuptake Inhibitors (SNRI)	From 1st dose to 4 weeks after the last dose

* No further differentiation if duration of impact is longer than 12 weeks since ICEs are considered in the analysis only for the first 12 weeks of the study.

In addition, the following concomitant medications are considered prohibited:

Table 6–6: Preliminary list of prohibited concomitant medication by drug names (can be updated during the blind review meeting)

WHO Drug Name	WHO Drug Record Number	WHO Drug Sequence Number 1	WHO Drug Sequence Number 2	Duration of impact with respect to efficacy
CLONIDINE	001711	01	001	From 1st dose to 4 weeks after the last dose
CLONIDINE HYDROCHLORIDE	001711	02	001	From 1st dose to 4 weeks after the last dose
DIXARIT	001711	02	002	From 1st dose to 4 weeks after the last dose
CLONIDINE HCL	001711	02	048	From 1st dose to 4 weeks after the last dose
CANNABIS SATIVA	002377	01	001	From 1st dose to 4 weeks after the last dose
MARIJUANA	002377	01	002	From 1st dose to 4 weeks after the last dose
CBD OEL	002377	08	003	From 1st dose to 4 weeks after the last dose
OXYBUTIN	005389	02	057	From 1st dose to 4 weeks after the last dose
DRIPTAN	005389	02	126	From 1st dose to 4 weeks after the last dose
GABAPENTIN	010030	01	001	From 1st dose to 4 weeks after the last dose
NEURONTIN [GABAPENTIN]	010030	01	002	From 1st dose to 4 weeks after the last dose

GABRION	010030	01	024	From 1st dose to 4 weeks after the last dose
GABAPENTINE	010030	01	045	From 1st dose to 4 weeks after the last dose
GABA [GABAPENTIN]	010030	01	226	From 1st dose to 4 weeks after the last dose
PREGABALIN	016141	01	001	From 1st dose to 4 weeks after the last dose
LYRICA	016141	01	002	From 1st dose to 4 weeks after the last dose
PRAGIOLA	016141	01	249	From 1st dose to 4 weeks after the last dose
EGZYSTA	016141	01	314	From 1st dose to 4 weeks after the last dose
PREATO	016141	01	722	From 1st dose to 4 weeks after the last dose
CANNABIDIOL	079492	01	001	From 1st dose to 4 weeks after the last dose

6.6 Appendix 6: Population characteristics

In general, descriptive statistics by treatment group and overall will be presented for variables defined in this section. For continuous variables, number of observations, mean, standard deviation, minimum, median, and maximum will be presented. For categorical variables, number and percentage of participants will be presented. Listings will be provided as appropriate.

6.6.1 Demographics and baseline characteristics

All demographic and baseline characteristics will be summarized by treatment group and overall. The descriptive statistics will be presented for the SAF and FAS.

Demographic and baseline assessments to be summarized will include:

- Age (at inclusion), region/country, race, ethnicity
- Categorized age
  - <40 years, 40-49 years, 50-59 years, 60-65 years, >65 years
- Weight (kg), height (cm), body mass index (BMI; kg/m^2)
- Categorized BMI (< 18.5, 18.5 to < 25, 25 to < 30, ≥ 30 kg/m^2)
- Smoking history (Never, Former, Current)
- Level of education

Demographic and baseline characteristics will be summarized also for the following subgroups:

- Region (North America, rest of the world)
- Race

- Ethnicity
- Smoking history (Never, Former, Current)
- BMI (< 18.5, 18.5 to < 25, 25 to < 30, ≥ 30 kg/m^2)

6.6.2 Reproductive and Menstrual History and History of menopause hormone therapy

Reproductive and menstrual history will include information on number of pregnancies, number of births, years being amenorrheic and number of participants with hysterectomy or oophorectomy. Hysterectomy or oophorectomy are based on Medical History. (For hysterectomy the PTs Hysterectomy, Hysterosalpingectomy, Hysterosalpingo-oophorectomy and Radical hysterectomy are considered. For oophorectomy the PTs Hysterosalpingo-oophorectomy, Oophorectomy, Oophorectomy bilateral, Salpingo-oophorectomy, Salpingo-oophorectomy bilateral, Salpingo-oophorectomy unilateral are considered.) These variables will be analyzed descriptively, separately for each treatment group and overall based on the SAF and FAS.

History of menopause hormone therapy will include information on history of menopause hormone therapy, contra-indications for hormonal treatment, personal risk factors for hormonal treatment and benefit risk assessment. This data will be shown in the listing only.

6.6.3 Protocol deviations

Important deviations from the protocol and validity findings and the resulting assignment of participants to the analysis sets (see Section 4) are agreed upon in the blind review meeting (BRM). The documentation of important deviations, validity findings and the assignment of participant data to analysis sets will be performed according to the sponsor's applicable Standard Operating Procedures and/or Instruction Manuals. The definition for important deviations and validity findings will be provided in the 'Specification of assessment criteria and identification requirements' before unblinding the data.

Identification of important deviations and validity findings will be done periodically while the study is running, concluding with the completion of the final list during the BRM. Any changes to the statistical analysis prompted by the results of BRM will be documented in an amendment and, if applicable, in a supplement to this SAP.

The list of important deviations and validity findings will only be final after assessment of conditional findings, which are validity findings that can only be identified after unblinding of study treatment.

6.6.4 Medical history

For medical history the MedDRA (current version at the time of analysis) will be used. Medical history findings (i.e., previous diagnoses, diseases or surgeries) not pertaining to the study indication, starting before start of treatment and considered relevant to the study will be tabulated by primary SOC and PT by treatment group and overall. Medical history will be presented for SAF.

All new or worsened findings after start of study treatment should be documented on the AE eCRF page.

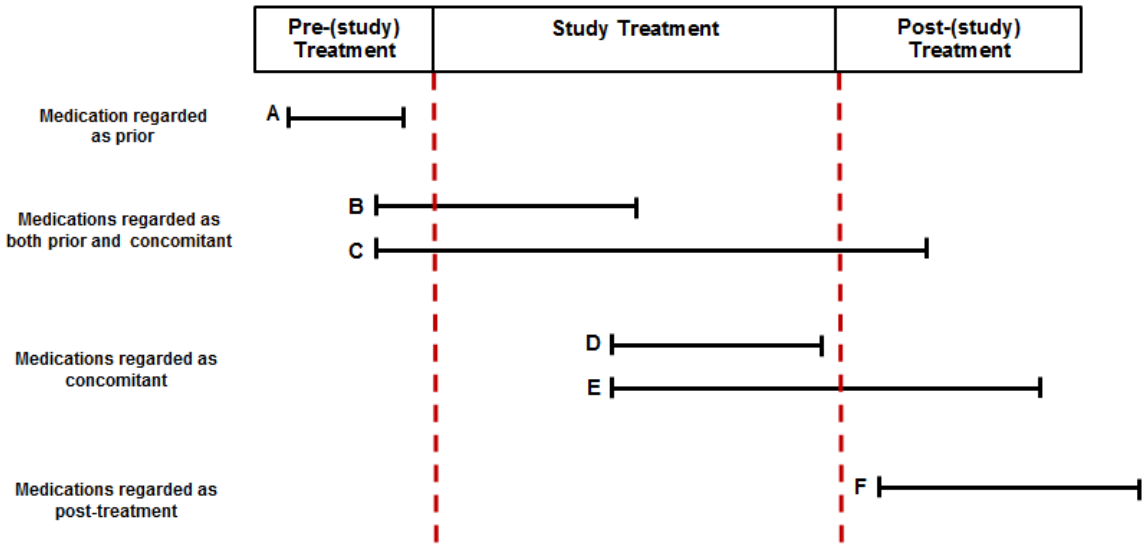
6.6.5 Prior, concomitant and post-treatment medication

For prior/concomitant/post-treatment medications, the following definitions in accordance with the Global Standards Catalogue (v4.0 or later) will be used in the analysis (see Figure 6–1):

- Prior medication: Medication taken before start of the study drug intake, (regardless of when it ended).
- Concomitant medication: Medication taken during treatment phase, i.e., between first and last study drug intake (regardless of when it started or ended).
- Post-treatment medication: Start of medication is after last study drug intake.

Missing dates will be imputed as described in Section 5.1.1.

Figure 6–1: Categories of concomitant medication (example).



Categories are prior medication (A, B, C), concomitant medication (B, C, D, E) and post-treatment medication (F). Source: Global Standards Catalogue V4.0

Medication, recorded as prior, concomitant or post-treatment medication in the eCRF, will be coded according to the World Health Organization Drug Dictionary WHODRUG Global (current version at the time of analysis), to the respective Drug Codes with their corresponding Anatomical Therapeutic Chemical (ATC) classification.

The number of participants taking prior, concomitant or post-treatment medication will be analyzed using frequency tables and shown by treatment group. Analysis of prior, concomitant and post-treatment medication will be done on the SAF.

6.7 Appendix 7: Coding conditions applicable for AESI

AESI	Search
Potential AESI – Liver event "Any condition triggering close liver observation" according to protocol Section 10.5 results in true AESIs of liver events. The search specified here is beyond the protocol definition of the AESI and will be considered together with the assessment by the Liver Safety Monitoring Board to determine a true AESI.)	SMQ Cholestasis and jaundice of hepatic organ SMQ Drug related hepatic disorders – severe events only SMQ Liver related investigations, signs and symptoms SMQ Liver-related coagulation and bleeding disturbances In addition, include AESIs ticked at AE-CRF.
Somnolence or fatigue	a) Somnolence MLG: PT Somnolence PT Hypersomnia b) Fatigue (MGL Decreased general strength and energy) PT Asthenia PT Decreased activity PT Fatigue PT Fatigue management PT Mental fatigue PT Physical deconditioning PT Sluggisheness In addition to MLGs above, include PT Sleep disorder due to general medical condition, hypersomnia type
Phototoxicity	BMQ Photosensitivity reactions
Post-menopausal uterine bleeding	MLG Female genital tract bleeding PT Abnormal uterine bleeding PT Abnormal withdrawal bleeding PT Cervix haematoma uterine PT Cervix haemorrhage uterine PT Coital bleeding PT Haematocoele female PT Haematosalpinx PT Haemorrhagic ovarian cyst PT Ovarian haematoma PT Ovarian haemorrhage PT Postmenopausal haemorrhage PT Unexpected vaginal bleeding on hormonal IUD PT Uterine haematoma PT Uterine haemorrhage PT Vaginal haematoma PT Vaginal haemorrhage PT Vulval haematoma PT Vulval haematoma evacuation PT Vaginal haemorrhage PT Withdrawal bleed

7. References

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

BIJUVA® Prescribing Information. FDA Prescribing information for BIJUVA® (estradiol and progesterone) capsules, for oral use. 2018.

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.

EMA. Guideline on clinical investigation of medicinal products for hormone replacement therapy of oestrogen deficiency symptoms in postmenopausal women. Committee for medicinal products for human use (CHMP). EMEA/CHMP/021/97 Rev. 1. 13 Oct 2005.

EMA. ICH guideline E17 on general principles for planning and design of multi-regional clinical trials. Committee for Medicinal Products for Human Use. EMA/CHMP/ICH/453276/2016 Rev.1. 14 Dec 2017.

FDA. Guidance for Industry. Drug-induced liver injury: premarketing clinical evaluation. July 2009.

FDA. Guidance for Industry. Estrogen and estrogen/progestin drug products to treat vasomotor symptoms and vulvar and vaginal atrophy symptoms – recommendations for clinical evaluation. Center for Drug Evaluation and Research (CDER), Draft Guidance. January 2003.

Guizzaro L, Pétavy F, Ristl R, Gallo C (2021) The use of a variable representing compliance improves accuracy of estimation of the effect of treatment allocation regardless of discontinuation in trials with incomplete follow-up. Statistics in Biopharmaceutical Research, 13:1, 119-127.

Hilditch JR, Lewis J, Peter A, van Maris B, Ross A, Franssen E, et al. A menopause-specific quality of life questionnaire: development and psychometric properties. Maturitas. 2008;61(1-2):107-21.

Lewis J, Hilditch JR, Wong C, Further psychometric property development of the Menopause-Specific Quality of Life questionnaire and development of a modified version, MENQOL-Intervention questionnaire. Maturitas. 2004; 50: 209–221

Little, R. (1993) Pattern-Mixture Model for Multivariate Incomplete Data. 126 Journal of the American Statistical Association (Biometrics, 1996, vol. 52, 1324-1333.), 1993.

NCT02865538. Evaluation of the Pharmacokinetics and Safety of BAY3427080 (NT-814) in Post-Menopausal Women With Vasomotor Symptoms (RELENT-1). Accessed: 30 Nov 2020.

NCT03596762. A Study of BAY3427080 (NT-814) in the Treatment of Moderate to Severe Post-menopausal Vasomotor Symptoms (SWITCH-1). Sponsored by Bayer. Accessed: 30 Nov 2020.

Pachman DR, Jones JM, Loprinzi CL. Management of menopause-associated vasomotor symptoms: Current treatment options, challenges and future directions. Int J Womens Health. 2010 Aug 9;2:123-35.

PREMARIN® Prescribing information. FDA Prescribing information for PREMARIN® (conjugated estrogens) Vaginal Cream. 2017.

PROMIS Sleep Disturbance – Scoring Manual
Available online:
https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Sleep_Disturbance_Scoring_Manual. Accessed: 24 Mar 2021.

Rance NE, Dacks PA, Mittelman-Smith MA, Romanovsky AA, Krajewski-Hall SJ. Modulation of body temperature and LH secretion by hypothalamic KNDy (kisspeptin, neurokinin B and dynorphin) neurons: a novel hypothesis on the mechanism of hot flushes. Front Neuroendocrinol. 2013 Aug;34(3):211-27.

Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharmaceut. Statist.2013,12 337–347.

Ratitch B, Lipkovich I, O'Kelly M. Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG 2013 - Paper SP0.

Rubin, DB. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987.

Stokes ME, Davis CS, Koch GG. Categorical Data Analysis Using SAS, Third Edition. SAS Institute, 2012.

Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS sleep disturbance and Sleep-Related Impairment item banks. Behav Sleep Med. 2011 Dec 28;10(1):6-24.

Title Page

Protocol Title: A double-blind, randomized, placebo-controlled multicenter study to investigate efficacy and safety of elinzanetant for the treatment of vasomotor symptoms over 26 weeks in postmenopausal women.

Protocol Number: 21652

Compound Number: BAY 3427080 / elinzanetant

Short Title: Overall **A**ssessment of efficacy and **S**afety of elinzanetant **I**n patients with vasomotor **S**ymptoms (OASIS-2)

Acronym: OASIS 2

Sponsor Name: Bayer Consumer Care AG

Legal Registered Address: Peter-Merian-Strasse 84, 4052 Basel, Switzerland

Regulatory Agency Identifier Number(s):

Registry	**ID**
EudraCT	2020-004855-34

Date: 26 OCT 2023
Version: 1.0

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.
The approval of the Statistical Analysis Plan is documented in a separate signature document.

Table of Contents

Table of Tables

1. Introduction

Study 21652 is one of several Phase 3 studies, aimed at characterizing the efficacy and safety of elinzanetant for the treatment of vasomotor symptoms related to menopause.

As described in further detail in the statistical analysis plan, important intercurrent events (ICEs) for this study are defined as temporary treatment interruption, permanent treatment discontinuation of randomized treatment and intake of prohibited concomitant medications having impact on efficacy. Permanent discontinuation due to adverse event (AE) or lack of efficacy may be handled differently in the analysis depending on estimand for participants who remained untreated/background therapy and for participants who initiate alternative vasomotor symptoms (VMS) treatment.

Drug groupings, individual drugs and manual review were used in the study to identify alternative vasomotor symptoms treatment and prohibited concomitant medications that were defined in the protocol to influence efficacy.

In the following, the lists of considered alternative vasomotor symptoms treatment and prohibited concomitant medication together with the pre-defined washout time period for their effect are documented prior to unblinding and the final database release.

2. Lists of alternative vasomotor symptoms treatment and prohibited concomitant medication

2.1 Alternative vasomotor symptoms treatment

Although more alternative treatment options may be available, only drug groupings/ drug names are listed in the following that were reported in OASIS 1-3 studies as of finalization of this document:

Table 2–1: Final list of alternative VMS treatment by drug grouping

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2	Duration of impact with respect to efficacy*
		5	Hormone replacement therapy	2	Oestrogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous or intradermal.
		5	Hormone replacement therapy	3	Progestogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous or intradermal.

1633	Drugs for psychiatric disorders	111	Antidepressants	113	Selective serotonin reuptake inhibitors (SSRI)	From 1st dose to 4 weeks after the last dose

*In this column, the terms "1st dose" and "last dose" both refer to the start date and stop date of certain doses of concomitant medication. If the dose was changed, a new start date and/or stop date were reported.

Table 2–2: Final list of alternative VMS treatment by drug names

WHO Drug Name	WHO Drug Record Number	WHO Drug Sequence Number 1	WHO Drug Sequence Number 2	Duration of impact with respect to efficacy*
OXYBUTIN	005389	02	057	From 1st dose to 4 weeks after the last dose

*In this column, the terms "1st dose" and "last dose" both refer to the start date and stop date of certain doses of concomitant medication. If the dose was changed, a new start date and/or stop date were reported.

2.2 Prohibited concomitant medication having impact on efficacy

All here listed drug groupings/ drug names were reported in OASIS 1-3 studies as of finalization of this document, although other medications may also be prohibited per the protocol.

Table 2–3: Final list of prohibited concomitant medication by drug grouping

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2	Duration of impact with respect to efficacy* (considered up to 12 weeks**)
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	109	GnRH agonists	From 1st dose to 12 weeks after the last dose
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	110	GnRH antagonists	From 1st dose to 12 weeks after the last dose
		5	Hormone replacement therapy	2	Oestrogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous

						or intradermal.
		5	Hormone replacement therapy	3	Progestogens	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral, sublingual or intrauterine, 3, From 1st dose to 12 weeks or longer if the route is subcutaneous or intradermal.
		772	Cancer therapies	738	Endocrine antineoplastic therapy	From 1st dose to 12 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	114	Monoamine oxidase (MAO) inhibitors, non-selective	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	115	Monoamine oxidase A (MAO-A) inhibitors	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	112	Non-selective monoamine reuptake inhibitors	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	113	Selective serotonin reuptake inhibitors (SSRI)	From 1st dose to 4 weeks after the last dose
45	Drugs interacting with CYP3A	240	CYP3A inducers	225	Moderate CYP3A inducers	From 1st dose to 4 weeks after the last dose. Considered not to have an influence on efficacy, if route is auricular(otic) or ophthalmic.
45	Drugs interacting with CYP3A	240	CYP3A inducers	265	Strong CYP3A inducers	From 1st dose to 4 weeks after the last dose
1633	Drugs for psychiatric disorders	111	Antidepressants	1830	Antidepressant Serotonin Norepinephrine Reuptake Inhibitors (SNRI)	From 1st dose to 4 weeks after the last dose

*In this column, the terms "1st dose" and "last dose" both refer to the start date and stop date of certain doses of concomitant medication. If the dose was changed, a new start date and/or stop date were reported.
** No further differentiation if duration of impact is longer than 12 weeks since ICEs are considered in the analysis

only for the first 12 weeks of the study.

In addition, the following concomitant medications are considered prohibited:

Table 2–4: Final list of prohibited concomitant medication by drug

Any additional condition, so that considered as prohibited and considered as intercurrent event	WHO Drug Name	WHO Drug Record Number	WHO Drug Sequence Number 1	WHO Drug Sequence Number 2	Duration of impact with respect to efficacy* **
If newly started or dose modified during the first 12 weeks of study period	ARMOUR THYROID	000537	01	011	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	NATURE THROID	000537	01	027	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LEVOTHYROXINE	000680	01	001	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	L-THYROXINE	000680	01	006	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	L-THYROXIN	000680	01	014	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LEVAXIN	000680	02	002	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	SYNTHROID	000680	02	005	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	EUTHYROX	000680	02	007	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	THYROXIN	000680	02	014	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LEVOTHYROXINE [LEVOTHYROXINE SODIUM]	000680	02	020	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	THYREX	000680	02	028	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	THYROHORMONE	000680	02	029	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	T4	000680	02	031	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LETROX	000680	02	046	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	UNITHROID	000680	02	052	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LEVOTIROXINA [LEVOTHYROXINE SODIUM]	000680	02	054	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	L-THYROXIN HENNING	000680	02	058	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	TIROXIN [LEVOTHYROXINE SODIUM]	000680	02	062	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	L THYROXIN	000680	02	070	From 1st dose to 12 weeks after the last dose
If newly started or dose	EUTIROX	000680	02	074	From 1st dose to 12 weeks

modified during the first 12 weeks of study period					after the last dose
If newly started or dose modified during the first 12 weeks of study period	LEVOTHYROXIN	000680	02	081	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	TIROSINT	000680	02	087	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	EUTHYROX N	000680	02	138	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LEVOTIROXIN	000680	02	145	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	L THYROX	000680	02	163	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	TIROSINT SOL	000680	02	219	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LIOTHYRONINE	001433	01	001	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	LIOTHYRONIN	001433	02	020	From 1st dose to 12 weeks after the last dose
	CLONIDINE	001711	01	001	From 1st dose to 4 weeks after the last dose
	CLONIDINE HYDROCHLORIDE	001711	02	001	From 1st dose to 4 weeks after the last dose
	DIXARIT	001711	02	002	From 1st dose to 4 weeks after the last dose
	CLONIDINE HCL	001711	02	048	From 1st dose to 4 weeks after the last dose
	CANNABIS SATIVA	002377	01	001	From 1st dose to 4 weeks after the last dose
	MARIJUANA	002377	01	002	From 1st dose to 4 weeks after the last dose
	CBD OEL	002377	08	003	From 1st dose to 4 weeks after the last dose
	OXYBUTIN	005389	02	057	From 1st dose to 4 weeks after the last dose
	DRIPTAN	005389	02	126	From 1st dose to 4 weeks after the last dose
	GABAPENTIN	010030	01	001	From 1st dose to 4 weeks after the last dose
	NEURONTIN [GABAPENTIN]	010030	01	002	From 1st dose to 4 weeks after the last dose
	GABRION	010030	01	024	From 1st dose to 4 weeks after the last dose
	GABAPENTINE	010030	01	045	From 1st dose to 4 weeks after the last dose
	GABA [GABAPENTIN]	010030	01	226	From 1st dose to 4 weeks after the last dose
	PREGABALIN	016141	01	001	From 1st dose to 4 weeks after the last dose
	LYRICA	016141	01	002	From 1st dose to 4 weeks after the last dose
	PRAGIOLA	016141	01	249	From 1st dose to 4 weeks after the last dose
	EGZYSTA	016141	01	314	From 1st dose to 4 weeks after the last dose
	PREATO	016141	01	722	From 1st dose to 4 weeks after the last dose
	CANNABIDIOL	079492	01	001	From 1st dose to 4 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	THYRONAJOD	109689	02	006	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12	LEVOTHYROXINE;POTASSIU	109689	03	001	From 1st dose to 12 weeks after the last dose

weeks of study period	M IODIDE				
If newly started or dose modified during the first 12 weeks of study period	LEVOTHYROXINE AND LIOTHYRONINE [LEVOTHYROXINE;LIOTHYRONINE]	131345	01	008	From 1st dose to 12 weeks after the last dose
If newly started or dose modified during the first 12 weeks of study period	NOVOTHYRAL	131345	03	007	From 1st dose to 12 weeks after the last dose

*In this column, the terms "1st dose" and "last dose" both refer to the start date and stop date of certain doses of concomitant medication. If the dose was changed, a new start date and/or stop date were reported.
** No further differentiation if duration of impact is longer than 12 weeks since ICEs are considered in the analysis only for the first 12 weeks of the study.