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 52 weeks in postmenopausal women
NCT Number:	NCT05030584
Document Date:	14-Feb-2024

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 52 weeks in postmenopausal women

Protocol Number: 21810

Compound Number: BAY 3427080 / elinzanetant

Short Title: Overall Assessment of efficacy and Safety of elinzanetant In patients with vasomotor Symptoms (OASIS-3)

Acronym: OASIS 3

Sponsor Name: Bayer Consumer Care AG

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

Regulatory Agency Identifier Number(s):

Registry	**ID**
EudraCT	2021-000059-38

Date: 14 FEB 2024
Version: 4.0

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 21810 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	31 OCT 2023	This amendment of the SAP includes correction of typos, clarifications of wording and analyses as well as adding 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 (Section 5.1) 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 (Section 5.1.2). 4) Data handling of PRO assessments for participants after premature treatment discontinuation and for assessments out-of-window in case of missing scheduled assessments (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 5.3.3 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 uses now 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 updated 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.
		11) A correction of the p-value that will be reported based on the non-parametric rank ANCOVA in the sensitivity analyses (see Section 5.3.2.1). 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.
		12) A frequency table has been added for actual closer liver observations	Existing AESI table include only potential close liver observation cases.

3.0	22 JAN 2024	This amendment of the SAP includes correction of typos, clarifications of wording and analyses as well as adding 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) Typo is corrected in the Table 5–3 in the Equation: α is the covariate effect for the interaction (xv), instead of (xt).	Typo noticed during review of this document.
		2) A correction for definition of derivation of post-treatment week was done in Section 5.1.2; "Up to Week 12" was removed from the sentence: The post-treatment weeks, i.e. data collected after premature discontinuation of treatment, are counted from last treatment date +1 onwards.	There can be post-treatment weeks after Week 12.
		3) Wording for biopsies shown at End of Treatment in Section 5.6.3.10 was clarified.	Clarified the text to better reflect the current TLF specification.
		4) Table 6–6 was updated. 5) Section 6.8 was added, including Table 6–7 and Table 6–8 to add list of used HRT or some other drugs affecting bone mineral density.	Updated based on blind data review. List for drugs affecting BMD was missing from earlier versions.
4.0	14 FEB 2024	This amendment of the SAP includes additions and modification of some analyses. Changes have been incorporated in following Sections:	Needed changes were identified during review of this document.
		1) Individual participant presentations in Section 5.6.3.4.1 were modified.	Figures and listing were modified according to advice from the pre-NDA meeting with FDA.
		2) Categorical analysis of T-score was added in Section 5.6.3.8.	T-score was added as an additional safety measure.

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 are associated with 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 VMS, 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 in these women.

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

It is anticipated that elinzanetant will be a relevant improvement compared to currently 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 serious side effects associated with hormonal therapy. It is also anticipated to result in greater efficacy and a better tolerability profile than anti-depressants used for treating VMS.

This is a phase 3 study to assess the efficacy and safety of elinzanetant for the treatment of VMS related to menopause.

The SAP describes the final analysis of the study. Table, figure and listing specifications are contained in a separate document.

1.1 Objectives and Endpoints and Estimands

Objectives and endpoints for the study are listed in Table 1–1.

Table 1–1: Objectives and endpoints

Objectives	Endpoints
Primary	
• To evaluate the efficacy of elinzanetant for the treatment of VMS associated with the menopause	**Primary endpoint** Mean change in frequency of moderate to severe HFs from baseline to Week 12 (assessed by HFDD) **Exploratory endpoints** • Mean change in frequency of moderate to severe HF from baseline over time (assessed by HFDD) • Mean change in severity of moderate to severe HF from baseline over time (assessed by HFDD) • Mean change in frequency of mild, moderate, and severe HF from baseline over time (assessed by HFDD)*

Objectives	Endpoints
Secondary	
• To evaluate the efficacy of elinzanetant on: ○ sleep quality ○ menopause related quality of life ○ weight and body composition in women being treated for relief of VMS associated with the menopause	**Secondary endpoints:** • Mean change in PROMIS SD SF 8b total score from baseline over time • Mean change in MENQOL total score from baseline over time **Exploratory endpoints:** • Mean change in MENQOL domain and single item scores from baseline over time* • Mean change in body weight from baseline over time • Mean change in body composition 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 participant rating of "quite a bit" or "very much" sleep disturbance due to HFs from baseline over time (assessed by HFDD)*
• To evaluate the safety of elinzanetant for the treatment of VMS associated with the menopause	**Other endpoints:** • Number of participants with non-benign endometrial biopsy results • Mean percentage change in BMD from baseline to Week 52 • Number of participants with TEAEs • Number of participants with abnormal laboratory parameters • Mean change in sleepiness score assessed by Sleepiness Scale at Week 1, Week 4, Week 12, Week 24, Week 36, Week 50 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)

HF = Hot Flash, HFDD=Hot Flash Daily Diary, MENQOL=Menopause Specific Quality of Life Scale, BMD= Bone mineral Density, PK = Pharmacokinetics, PROMIS SD SF 8b=Patient-reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b, TEAE = treatment emergent adverse event, VMS = Vasomotor symptoms

*Additional exploratory endpoints

Estimands

The attributes of the estimand for the primary endpoint, i.e. mean change in frequency of moderate to severe HF from baseline to Week 12 (assessed by HFDD), are as follows:

- Population: Post-menopausal women aged 40-65 with VMS as described by the inclusion/exclusion criteria detailed in the protocol.
- Variable: Change in frequency of moderate to severe HF from baseline to Week 12
- Treatment: 120 mg elinzanetant, placebo
- Intercurrent events (ICEs). See Table 1–2. (For further details regarding the identification of ICEs see Section 6.5.)

- Population level summary: Mean change in frequency of moderate to severe HF from baseline to Week 12. Treatment comparison will be based on differences in treatment group means.

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, VMS = Vasomotor symptoms

*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 (days 15-28).

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 (days 43-56).

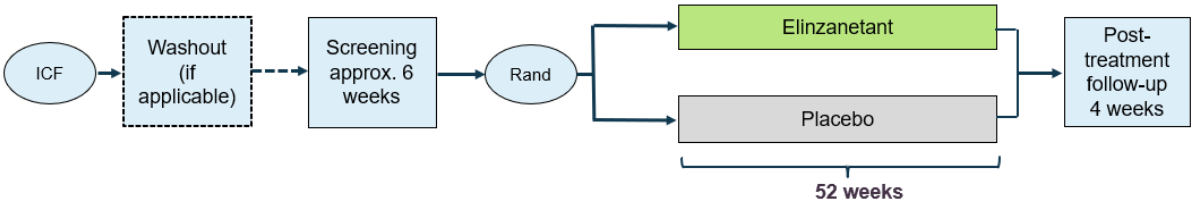
Week 12 = Treatment taken <80% during weeks 1-12 (day 1-84) OR treatment taken on <5/7 days during either week 11 or 12 (days 70-84).

**ICEs will be reviewed prior to the study unblinding

1.2 Study Design

An overview of the study design is shown in Figure 1–1.

Figure 1–1: Study schema



- This is a multi-center, multi-country, double-blind, randomized, parallel-group, placebo-controlled, Phase 3 intervention study in post-menopausal women with VMS.
- Primary efficacy endpoint will be evaluated at Week 12.
- Total study duration for all participants: approximately 62 weeks (plus potential washout period), including
  - Washout (if applicable)
  - Screening approximately 6 weeks
  - Treatment: 52 weeks
  - Follow-up: 4 weeks
- Eligible participants will be randomized 1:1 into two arms as follows:
  - elinzanetant
  - placebo
- Randomization will be stratified by region: North America, rest of the world (ROW)
- No interim analysis is planned.
- Primary analysis will be performed after database release after last patient last visit (LPLV).

2. Statistical Hypotheses

H_0: $\mu_P \leq \mu_V$ versus H_1: $\mu_P > \mu_V$ where μ_P and μ_V stand for the mean change in frequency of moderate to severe HF from baseline in the placebo (P) and verum (V) group to Week 12.

The type I error rate will be controlled at a one-sided α=0.025 level.

All other endpoints will be summarized descriptively and therefore no multiplicity adjustment is needed.

3. Sample Size Determination

Approximately 600 participants will be randomly assigned to study intervention. Of these, 300 participants will be randomly assigned to elinzanetant and the other 300 participants will be randomly assigned to the placebo arm. Assuming a yearly drop-out rate of 30 % and 10% during the months 1-3 and 4-6 in the elinzanetant arm, approximately 210 participants will be available that were treated for 1 year and approximately 240 participants will be available that were treated for at least 6 months.

The number of participants needed for this study is based on the total number of participants needed for the safety evaluation. Following the ICH E1 guideline (EMA 1995), a total of 1500 participants are required that were exposed to at least one dose of elinzanetant 120 mg. Of these 1500 participants 300-600 should have been treated for at least 6 months and approximately 100 for 1 year. Based on the number of participants that are available from previous phase 1 and phase 2 studies, together with the number of participants that are planned to be enrolled in upcoming phase 3 studies, the number of participants needed for the elinzanetant arm in this study is 300, also taking into account the assumed drop-out described above.

Assuming a screen failure rate of 30% approximately 860 participants need to be enrolled to achieve the required number of 600 randomized participants. Additionally, further participants may be enrolled in case of an exceptionally high drop-out rate due to a trial-continuity issue.

A formal sample size justification was performed for the primary efficacy endpoint using the one-sided two-Sample T-test (equal variance) assuming at least approximate normal distribution. Drop-out rate during the first 3 months is assumed to be 10%. The power with treatment differences (i.e. mean change in frequency of moderate to severe HFs from baseline to Week 12), with N=270 per treatment group and assumed standard deviation are provided in Table 3–1. The data from placebo arm in SWITCH-1 study (NCT03596762, 2020) was used for assumption of standard deviation.

Table 3–1: Sample size justification

Standard deviation	Treatment difference elinzanetant vs. placebo	Power %
4.29	-1.00	77.1 %
	-1.50	98.2 %
	-2.00	>99.9 %

PASS version 13.0.11 was used for sample size calculation.

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.
BMD Analysis Set (BAS)	All randomized participants in sites assigned to perform DEXA, who have baseline and at least one post-baseline (week 24 and/or week 52) DEXA scan available and have not used HRT or some other drugs affecting bone density (other than vitamin D and calcium) until that time point

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 participants will be analyzed according to the intervention received. The analysis of the Bone Mineral Density (BMD) will be based on the BAS and participants will be analyzed according to 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

All variables collected in the study will be summarized using descriptive statistical methods. Continuous variables will be summarized using at least the following descriptive statistics: number of non-missing observations, arithmetic mean and standard deviation (SD), median, minimum and maximum. The geometric mean and geometric SD will be provided instead of the arithmetic mean and SD for variables where lognormal distributions are assumed. Categorical data will be summarized using frequency tables. Where appropriate, the data will also be presented by visit or week, including the analysis of the changes from baseline.

For the main analysis the type I error rate will be controlled at 2.5 % one-sided. Where appropriate two-sided 95 % confidence intervals (CIs) will be provided.

The statistical evaluation will be performed by using the software SAS (release 9.4 or higher; SAS Institute Inc., Cary, NC, USA).

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

Number of decimal places for summary statistics are shown in Table 5–1. Derived variables would have maximum 1 decimal place and that value is considered as original data.

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

Statistic	Number of digits
Minimum, maximum	Same as original data or maximum 3 decimal points
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 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 5.3.1 and 5.3.3. 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 Case Report Form (CRF) or electronic diary (eDiary).

For the computation of durations, i.e., time between start and end date 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.

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 respondent missing an entire PRO assessment for a given time point (e.g. a given day (i.e. morning and evening eDiary) 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 per se in so far that respondents have to select an answer for an item in order to move on to the next item on the eDiary/electronic hand-held device. This does however 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 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 5.1.2.1). 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 5.3.1.

5.1.1.1.2 Patient-reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b (PROMIS SD 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. Missing data will not be imputed.

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 scores 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 items 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. 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 are described in Section 6.2.3.

5.1.1.1.4 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].

Derivation of intercurrent event (ICE): Intercurrent events that 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 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 or 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 5.3.1.

Derivation of post-treatment weeks for HFDD: The post-treatment weeks, i.e. data collected after premature discontinuation of treatment, 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 post-treatment 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 post-treatment week 1 will include the data from days 40-46. Data from post-treatment weeks are utilized in the model-based analyses, and shown in the listings, but not included in the descriptive tables.

Laboratory values <X or >X: For laboratory values lower than a limit of detection, which are given as <X, half the value of the X will be used for analysis (i.e., values reported as "<X", "<X.X"," <X.XX" or etc. the value for analysis used will be derived by "X/2", "X.X/2"," X.XX/2" etc.). Differences between 2 values <X will be assigned values of 0. Ratios between 2 values <X will be assigned a value of 1. For values which are higher than a limit of detection Y (given 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 valid 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 of PROMIS SD SF 8b and MENQOL 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	1	8-9
PROMIS SD SF 8b	2	15-16
PROMIS SD SF 8b	3	22-23
PROMIS SD SF 8b, MENQOL	4	29-30
PROMIS SD SF 8b, MENQOL	8	55-57
PROMIS SD SF 8b, MENQOL	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): 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 18, Week 24, Week 36 and Week 52.

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.
- Assessments from follow-up visits will be shown as follow-up in the descriptive tables.
- In the descriptive tables, data collected after premature discontinuation of treatment 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.

5.1.2.1 Definition of HFDD related 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 12/over time:

This will be calculated as the difference in the mean daily frequency of moderate to severe HF at Week 12 or at various weeks, 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 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 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. 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 over time:

This will be calculated as the difference in the mean daily severity of HF at various weeks, from the baseline value of the severity of moderate to severe HF.

5.2 Participant Dispositions

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

Other summary statistics will include:

- The number of screening failures and the reason for failed screening
- The number of participants randomized but not treated (overall and by treatment group).
- The 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, follow-up, 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 subject 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 Endpoint Analysis

The primary endpoint of this study is:
Mean change in frequency of moderate to severe HFs from baseline to Week 12. See definition in Section 5.1.2.1.

5.3.1 Main Analytical Approach

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

The frequency 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 weekly until Week 12 and at Weeks 23, 24, 35, 36, 49 and 50 thereafter. 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.

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 (from least restrictive to most restrictive): 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 be applied for each imputed dataset.

Equation	$Y_{ijkl} = \mu + \beta x_i + t_k + r_l + v_j + (tv)_{kj} + \alpha(xv)_{ij} + s_i + \varepsilon_{ijkl}$ 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 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 ICEs in this trial in the main estimand. According to this policy, 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) 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 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 multiple imputation (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 (Guizzaro et 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 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 as the number of participants, estimated mean change from baseline and standard error (SE) for each treatment arm at Week 12 as well as the estimated treatment difference (elinzanetant - placebo), SE of the difference, associated 95% CI and P-value (one-sided).

5.3.2 Sensitivity Analyses

5.3.2.1 Assessment of normality assumption

The assumption for normality in the main analysis (see Section 5.3.1) 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 (Stokes et al. 2012). This will be carried out for the primary endpoint; change from baseline of HF frequency 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.

5.3.2.2 Tipping point analysis

A tipping point analysis will be applied to assess the sensitivity of the main analysis results to modeling of the missing and unobservable data that occur in presence of ICEs as described in Section 5.3.1. 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 endpoint 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. Additional details on the tipping point sensitivity analysis are provided in Section 6.4.

5.3.3 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.

5.3.3.1 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 medication having impact on efficacy	All reasons	**Hypothetical,** If participant did not take prohibited concomitant medication.	Model outcomes under the hypothetical scenario during the period of 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 (see Sections 6.5.1 and 6.5.3 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 in 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* 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.	
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	**Treatment policy/Hypothetical**	Missing values that occur after the ICE will be

	discontinuation Due to AE/lack of efficacy	**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.	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 Section 1.1), 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 the Rubin's rule (Rubin, 1987) to obtain an overall estimated treatment effect.

The same baseline covariates will be included in the imputation and the analysis step as specified in Section 5.3.1.

The results from the supplementary analysis will be presented in a similar way as to the main analysis (see Section 5.3.1).

5.3.3.2 Second Supplementary Analysis

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 5.3.3.1).

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: 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.
- 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.

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.

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**
NA	Intake of prohibited medications Any reason	**Composite strategy** Treatment failure (i.e., no improvement from baseline.) Discard the collected data after the intake of prohibited medication(s) for a certain time	N/A

		frame based in the medication class (See Section 6.5.3 for details) and assign the value of zero change from baseline.	
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 utilize the collected data.	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). **(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 impute with participant's baseline value.	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 medication).

5.4 Secondary Endpoints Analysis

The secondary efficacy endpoints in this study are:

- Mean change in PROMIS SD SF 8b total score from baseline over time
- Mean change in MENQOL total score from baseline over time

The calculation of the scores for PROMIS SD SF 8b and for the MENQOL total score will be done according to the respective questionnaire guidelines.

5.4.1 Definition of Endpoints

5.4.1.1 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 over time

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 to 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.

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

- Mean change in MENQOL total score from baseline over time

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 on Section 6.2.3.

5.4.2 Main Analytical Approach

PROMIS SF 8b total raw scores, converted T-scores and MENQOL total scores will be summarized descriptively by treatment group and week 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 in the PROMIS SF 8b T-scores and MENQOL total scores will be shown using line plots 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 12 and 52 will be presented for the MENQOL total score and for the PROMIS SD SF 8b T-score.

In addition, PROMIS SD SF 8b raw score for each item will be summarized by a frequency table with the number of observation and percentage by treatment group and by week. Bar charts will also be created by treatment group over time.

For possible imputation of missing values, see section 5.1.1.1.

5.5 Exploratory Endpoints Analysis

The following exploratory endpoints will be analyzed by descriptive statistical methods as described in Section 5.1 by treatment group. The analysis will be based on the FAS. The ICEs will be handled according to the treatment policy. No missing data imputation is planned for the exploratory endpoints.

5.5.1 HFDD related exploratory endpoints

- Mean change in frequency of moderate to severe HF from baseline over time (assessed by HFDD)
- Mean change in severity of moderate to severe HF from baseline over time (assessed by HFDD)
- Mean change in frequency of mild, moderate, and severe HF from baseline over time (assessed by HFDD)
- Mean change in frequency of nighttime awakening from baseline over time (assessed by HFDD)
- Mean change in proportion of days with participant rating of ‘quite a bit’ or ‘very much’ sleep disturbances experienced due to HF from baseline over time (assessed by HFDD).

The mean change in frequency of mild, moderate, and severe HF from baseline over time will be calculated similar to the mean change in frequency of moderate to severe HF described in Section 5.1.2.1, 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 not available 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.

Calculation of mean change in severity of moderate to severe HF has been described in section 5.1.2.1. Summary statistics for the change from baseline in severity of moderate to severe HF and change from baseline in the frequency of mild, moderate, and severe HF will be summarized similar to change in frequency of moderate to severe HFs presented in Section 5.3.1. Relative change (%) from baseline will be additionally displayed only for the severity of moderate to severe HFs. The change from baseline, relative change from baseline in severity of moderate to severe HF over time and the change from baseline in the frequency of mild, moderate, and severe HFs over time will be shown using line plots for means together with 95% CIs by treatment group.

A graphical summary with relative 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 reduction on the y-axis, by treatment group, for week 12 and 52 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 change on the y-axis, by treatment group, for week 12 and 52 will also be produced.

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 (morning diary). 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.

Summary statistics in frequency of nighttime awakenings as well as the absolute change from baseline will be presented by weekly until Week 12 and at Weeks 23, 24, 35, 36, 49 and 50 thereafter 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. A graphical summary 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 12 and 52 will also be provided.

The calculation using HFDD 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.

Sleep disturbances due to HF are assessed every morning as part of the HFDD (morning diary). 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 participants having reported '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 participants having reported 'quite a bit' or 'very much' 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 by treatment group will be shown using line plots for means together with 95% CIs by treatment group.

5.5.2 MENQOL related endpoint

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

The MENQOL domain scores and single item scores will be analyzed similar to MENQOL total score, as described in Section 5.4.2. Converted scores will be shown in the tables.

5.5.3 Body composition related endpoints

- Mean change in body weight from baseline over time
- Mean change in body composition from baseline over time (fat %, fat free mass (FFM), muscle mass, total body water (TBW), bone mass and visceral fat rating)

Weight and body composition measures will be summarized using descriptive statistics by treatment group for each week, as well as absolute changes from baseline to various weeks will be presented. The change from baseline over time in the weight and body composition measures will be shown using line plots for means together with 95% CIs by treatment group.

5.6 Safety Analyses

5.6.1 Extent of Exposure

The analyses described in this section will be presented for the SAF, the FAS and the BAS. All measures will be presented for weeks 1-12, weeks 13-52 and overall.

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.

The extent of exposure to elinzanetant will be summarized as the total amount of study drug intake in grams and the 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, percent 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.

Treatment duration and 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. For eDiary based presentation, Weeks 1-12 will be defined as days 1-84 and Weeks 13-52 as days from 85 onwards until last study treatment intake. For eCRF page approach, Weeks 1-12 will be defined from start of study treatment until (T3 visit date -1) and Weeks 13-52 as from T3 visit date until end of study treatment intake. 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.

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 by treatment group 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, treatment-emergent (TEAE) and post-treatment AEs. Incidences of TEAEs and TESAEs per 100 person-years will be summarized by treatment group. The rate per 100 person-years is calculated as

Rate per 100 person-years = number of subjects / (total drug exposure in years / 100),

where 365.25 days are taken as one year.

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 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). 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 also 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 (with regression lines) showing blood pressure values at Week 12 and 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 graphically displayed with boxplots. Line plots depicting the mean change from baseline ± 95% CI will be provided for CK, LDH, estradiol, HbA1c, total cholesterol, HDL, LDL, 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 the 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 (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 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 participant 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 participants 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 participants 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 participant presentations

If a participant met close liver observation criteria, 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), in addition all AEs and CM 6 months (182 days) prior to the first onset of close liver observation will be shown in the plot 1. Furthermore, listing will be provided for INR

(absolute values) and for liver-related parameters, i.e. ALT, AST, Total and direct bilirubin, ALP, GGT, CK and LDH (with results relative to ULN).

The liver injury criteria

All subjects meeting close liver observation criteria will be assessed by Liver Safety Monitoring Board (LSMB). Number of subjects fulfilling the liver injury criteria per LSMB manual will be presented with a frequency table.

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), at week 52 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, Week 52 and EoT) will be presented by treatment group.

Endometrial thickness will be summarized by treatment group for each visit, including change from baseline where appropriate, 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 Bone mineral density

Bone mineral density (BMD) will be analyzed using BMD Analysis Set (BAS). List of HRT and other drugs affecting bone density will be given in the Section 6.8.

Summary statistics for the percentage change from baseline in BMD of lumbar spine, total hip and femoral neck (using the central reading results) will be provided by treatment group and visit.

The time course by visit and by treatment group will be shown using line plots depicting the mean ± 95% plot for the percentage change from baseline in BMD of lumbar spine, total hip and femoral neck.

Percentage change in BMD of lumbar spine, total hip and femoral neck from baseline with frequencies and percentages for the following categories: <-6%, ≥-6% - < -3% , ≥-3% - < -1.5% , ≥-1.5% - ≤0%, >0% - ≤1.5%, >1.5% - ≤3% , >3% - ≤6% and >6% will also be summarized by treatment group and visit.

The Z-score and the T-score and the respective change from baseline in Z-score and T-score by BMD location will be summarized by treatment group for each visit using descriptive statistics.

Z-score and T-score with frequencies and percentages for the following categories: ≤-2, >-2 - ≤-1.5 , >-1.5 - ≤-1 , >-1 - ≤0, >0 - ≤1, >1 - ≤1.5, >1.5 - ≤ 2 and >2 will also be summarized by treatment group and visit.

Participants that had at least one DEXA scan with either BMD loss >6% from baseline to Week 24/Week 52 for any location or Z-score ≤ -2 SD for any location will be listed.

If the baseline scan at Screening Visit 2 is missing or not evaluable based on the evaluation from the imaging central laboratory, the participant is requested to have a repeat scan. If it's not possible to have this repeat scan performed before the start of treatment, this repeat scan is allowed to be performed shortly after the start of the treatment (within maximum of 15 days). Considering the BMD will not change quickly, this can be accepted from a medical point of view. If a BMD value performed after the treatment start date has been flagged as baseline, it will be analyzed as baseline.

Both scheduled and unscheduled measurements can be used in the analysis. The following time windows apply, when re-mapping measurements to the visits:

- Week 24: Day 90 after Start of Treatment (SoT) to Day 239 after SoT , but no later than the date of last dose + 60 days
- Week 52: Day 270 after SoT to Day 419 after SoT, but no later than the date of last dose + 60 days

Correlation analyses

To evaluate relationship between percentage change in BMD of lumbar spine, total hip and femoral neck from baseline to Week 24 and to Week 52, and demographic variables (e.g., age, weight, BMI, BMD at baseline) scatter plots will be created.

To study effect of race to change in BMD of lumbar spine, total hip and femoral neck, distribution of percentage change in BMD from baseline to Week 24 and to Week 52 by race will be graphically depicted with boxplots.

5.6.3.9 Cervical cytology

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

5.6.3.10 Endometrial biopsy

Analyses will present the number and percentage of either participants or biopsies for women without hysterectomy. Biopsies taken at End of Treatment (including biopsies taken at Week 52 or at earlier EoT visit) or at Unscheduled visit will be further divided in the tables for subjects treated at least 326 days and for subjects treated less than 326 days.

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"

5.6.3.11 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, 12, 24, 36 and 50.

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 single item scores (i.e., morning, afternoon and evening scores).

Summary statistics including corresponding change from baseline 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 of 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.

5.6.3.12 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 and by visit 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 pharmacokinetic (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 data will be described in a separate document and the results will be presented outside Clinical Study Report (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 Pharmacodynamics

The analyses for pharmacodynamics (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 will be used to assess psychometric properties of scores form HFDD, PROMIS SD SF 8b and the MENQOL 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 the 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)
- Moderate to severe HFs at baseline (< 35, ≥ 35)

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

Descriptive statistics for PROMIS SD SF 8b T-scores (converted from raw scores) will be provided for the following subgroups based on Insomnia Severity Index (ISI) administered at baseline (see Section 6.6.6 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 event 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
BL	Baseline
BMI	Body mass index
BMD	Bone mineral density
BRM	Blind review meeting
CI	Confidence interval
CK	Creatinine kinase
CM	Concomitant medications
CMH	Cochran-Mantel-Haenszel
COVID-19	Coronavirus disease of 2019
CS	Compound symmetry
CSR	Clinical study report
DEXA scan	Dual energy x-ray absorptiometric scan
DILI	Drug-induced liver injury
eC-SSRS	Electronic Columbia-Suicide Severity Rating Scale
ECG	Electrocardiogram
eCRF(s)	Electronic case report form(s)
eDiary	Electronic diary
EIN	Endometrioid Intraepithelial Neoplasia
EMA	European Medicines Agency
EoT	End of treatment
ePROs	Electronic Participant-Reported Outcomes
FAS	Full analysis set
FDA	US Food and Drug Administration
GnRH	Gonadotropin-releasing hormone
GGT	Gamma-glutamyl transferase
HF(s)	Hot flash(es)
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
LDH	Lactate dehydrogenase
LPLV	Last patient last visit
LSMB	Liver Safety Monitoring Board
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
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
SoT	Start of treatment
SSRI	Selective serotonin reuptake inhibitors
TB	Total bilirubin
TEAE(s)	Treatment-emergent adverse event(s)
UN	Unstructured
ULN	Upper limit normal
VMS	Vasomotor symptoms

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

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 from Visit SCR-2 until Visit T3, and during Week 23 & 24, Week 35 & 36 and Week 49 & 50. 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 is defined as a "sensation of heat without sweating", moderate HF is defined as a "sensation of heat with sweating, but able to continue activity", and severe HF is 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).

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, 18, 24, 36, 52 and 56 (i.e. Follow-up).

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 total MENQOL score the aggregated mean of the mean domain scores is calculated. Higher scores indicate greater bother. Total MENQOL score and domain scores and individual item scores are reported 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, 24, 36, 52 and 56 (i.e., Follow-up).

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

A pattern mixture modeling approach (Little, 1993) will be used in combination with multiple imputation (Ratitch B, O'Kelly M; Guizzarro 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 for frequency 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 = 21810" 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. 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 pattern 1 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 (Guizzarro 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 5.3.1).

   The estimated treatment difference at Week 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 primary endpoint, i.e., frequency of HF:

1. The imputation step will be done as described in Section 5.3.1.
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 5.3.2.2 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 5.3.1.
4. Steps 1-3 will be repeated with increasing settings of "delta" (see Section 5.3.2.2) for primary endpoint until the estimated treatment difference at 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.

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* (considered up to 12 weeks**)
		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.
** 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.

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* (considered up to 12 weeks**)
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.
** 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.

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/names were reported in OASIS 1-3 studies, 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

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.

The list of prohibited concomitant medication with the pre-defined washout 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 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. 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 6–6: Preliminary list of prohibited concomitant medication by drug names (can be updated during the blind review meeting)

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* (considered up to 12 weeks**)
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	UNITHROID	000680	02	052	From 1st dose to 12 weeks after the last dose

12 weeks of study period					
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 modified during the first 12 weeks of study period	EUTIROX	000680	02	074	From 1st dose to 12 weeks 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
	CANNABIS SATIVA	002377	01	002	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 weeks of study period	LEVOTHYROXINE;POTASSIUM IODIDE	109689	03	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	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.

6.6 Appendix 6: Population characteristics

In general, descriptive statistics 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, FAS and BAS.

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 (only for FAS):

- 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)
- Moderate to severe HFs at baseline (< 35, ≥ 35)

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, FAS and BAS.

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 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.

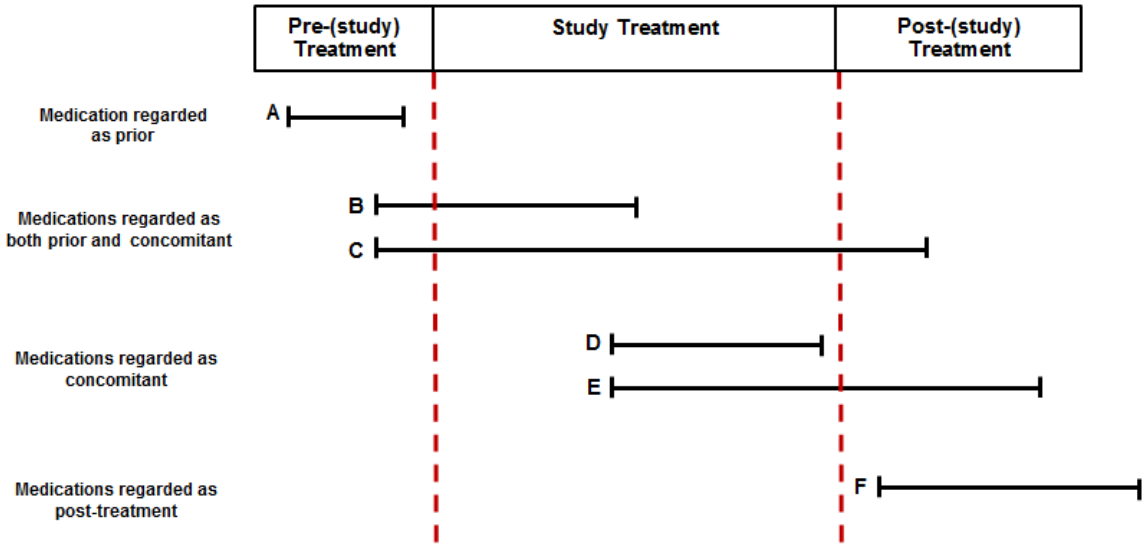
6.6.5 Prior, concomitant and post-treatment medication

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 and overall. Analysis of prior, concomitant and post-treatment medication will be done on the SAF.

6.6.6 Insomnia Severity Index (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:

1. difficulty falling asleep
2. difficulty staying asleep
3. problems waking up too early
4. satisfaction/dissatisfaction with current sleep pattern
5. noticeability of sleep problems by others
6. worries/distress caused by the sleep difficulties
7. 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)

The absolute values for Total score will be analyzed at baseline by descriptive statistical methods as described in section 5.1 by treatment group and overall. Summary statistics of Total score across the different severity categories by treatment group will be provided with frequency tables as well. In addition, Severity of insomnia presented in categories above will be summarized with frequency tables by treatment group and overall.

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

6.8 Appendix 8: List of HRT and other drugs affecting bone mineral density

Manual review will be used to identify HRT and other drugs affecting bone mineral density based on the Table 6–7 and Table 6–8. Participants, who used any drugs affecting bone during the study will be excluded from the BMD analysis set. The list of the drugs affecting bone will be reviewed during the blind review of the data and finalized prior to unblinding.

Table 6–7: List of HRT and other drugs affecting bone mineral density

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	109	GnRH agonists
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	110	GnRH antagonists
		99	Drugs used in diabetes	106	Thiazolidinediones
		5	Hormone replacement therapy	2	Oestrogens
		5	Hormone replacement therapy	3	Progestogens
274	Immunomodulators	186	Immunosuppressant drugs	418	Calcineurin inhibitors
126	Antithrombotic drugs	523	Heparins	524	Low molecular weight heparins (LMWH) and heparinoids
126	Antithrombotic drugs	523	Heparins	525	Other heparins
		772	Cancer therapies	738	Endocrine antineoplastic therapy
1633	Drugs for psychiatric disorders	111	Antidepressants	113	Selective serotonin reuptake inhibitors (SSRI)

Table 6–8: List of HRT and other drugs affecting bone mineral density by drug code

Drug group/class	Example	Drug Code	Reason	Comments
Bisphosphonates	alendronate risedronate zoledronic acid ibandronate	90049001001	treatment to Osteoporosis	ATC M05BA, Bisphosphonates
RANK ligand inhibitors	Denosumab(Prolia, or Xgeva)	90121401001	treatment to Osteoporosis	ATC M05BX, Other drugs affecting bone structure and mineralization and/or SDG subgroup Monoclonal antibodies - non antineoplastic
parathyroid hormone and parathyroid hormone–related protein analogs	Teriparatide abaloparatide	90089101001	treatment to Osteoporosis	ATC H05A Parathyroid hormones and analogues and/or SDG subgroup Other hormones and hormone modulators used as immunomodulators
Selective estrogen receptor modulators	raloxifene, bazedoxifene	90088001001	treatment to Osteoporosis	ATC G03XC SELECTIVE ESTROGEN RECEPTOR MODULATORS and/or SDG subgroup Oestrogens used in replacement therapy
Calcitonin	Calcitonin (Salmon)	90028101001	treatment to Osteoporosis	ATC H05BA Calcitonin preparations and/or SDG subgroup Other hormones and hormone modulators used as immunomodulators
Glucocorticoids	dexamethasone	90039201001	induce osteoporosis (side effect)	ATC H02AB,R03BA Glucocorticoids and/or SDG Corticosteroids
Antiepileptic drugs	phenytoin phenobarbital carbamazepine valproate	90035201001	induce osteoporosis (side effect)	ATC N03A Antiepileptics and/or SDG Anticonvulsants

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.

EMA. ICH Harmonised Tripartite Guideline: The extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life-threatening conditions. ICH Topic E 1. CPMP/ICH/375/95. June 1995.

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

Guizzaro Lorenzo, Pétavy Frank, Ristl Robin & Gallo Ciro (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. Pattern-Mixture Model for Multivariate Incomplete Data. 126 Journal of the American Statistical Association. Vol. 88, No. 421 (Mar., 1993), pp. 125-134

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. Pharm Stat. 2013 Nov-Dec;12(6):337-47

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.