STATISTICAL ANALYSIS PLAN

Sponsor	**Helsinn Healthcare SA** Via Pian Scairolo 9 6912 Lugano, Switzerland Phone: +41 91 985 21 21 Fax: +41 91 985 21 22
Clinical Trial Protocol Identification No.	ANAM-17-20
Title:	A phase 3, randomized, double-blind, placebo-controlled, multicenter study to evaluate the efficacy and safety of anamorelin HCl for the treatment of malignancy associated weight loss and anorexia in adult patients with advanced non-small cell lung cancer (NSCLC).

Trial Phase	Phase 3
Investigational Medicinal Product(s)	Anamorelin HCl
Clinical Trial Protocol Version	Final (v5.A/5.0) – 13-Jul-2022
Statistical Analysis Plan Author	[redacted], Biostatistician, [redacted], Biostatistician Business & Decision Life Sciences [redacted], Senior Biostatistician Consultant Helsinn Healthcare SA
Statistical Analysis Plan Date and Version	12 May 2023 Final v1.0

SIGNATURES PAGE

Statistical Analysis Plan - Protocol number: ANAM-17-20
Title of the protocol:
A phase 3, randomized, double-blind, placebo-controlled, multicenter study to evaluate the efficacy and safety of anamorelin HCl for the treatment of malignancy associated weight loss and anorexia in adult patients with advanced non-small cell lung cancer (NSCLC).

Sponsor's biostatistician	Helsinn Healthcare SA
Name: [redacted]	Via Pian Scairolo 9
Function: Senior Biostatistician Consultant	6912 Lugano, Switzerland

Signature & Date: [redacted]

Sponsor's Clinical Operation Manager	Helsinn Healthcare SA
Name: [redacted]	Via Pian Scairolo 9
Function: Manager, Clinical Operations	6912 Lugano, Switzerland

Signature & Date: [redacted]

B&D author of the SAP	Business & Decision Life Sciences
Name: [redacted]	Rue Saint Lambert 141
Function: CRO Biostatistician	1200 Brussels, Belgium

Signature & Date: [redacted]

TABLE OF CONTENTS

1. LIST OF ABBREVIATIONS

5-IASS	5-item Anorexia Symptom Subscale score
ACS	Anorexia/Cachexia Subscale
ADaM	Analysis Data Model
AE	Adverse Event
AESI	Adverse Event of specific interest
ALT	Alanine Aminotransferase
ANAM	Anamorelin
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AST	Aspartate Aminotransferase
BDRM	Blinded Data Review Meeting
BL	Baseline
BMI	Body Mass Index
BP	Blood Pressure
CDISC	Clinical Data Interchange Standards Consortium
CCR	Composite clinical response
CI	Confidence Interval
CRF	Case Report Form
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
ePRO	Electronic Patient Reported Outcome
EOS	End of Study
EOT	End of Treatment
FAACT	Functional Assessment of Anorexia/Cachexia Treatment
FACIT	Functional Assessment of Chronic Illness Therapy
FACIT-F	Functional Assessment of Chronic Illness Therapy-Fatigue
FACIT-G	Functional Assessment of Cancer Therapy - General
FAS	Full Analysis Set
FDA	Food and Drug Administration
FU	Follow up
GCP	Good Clinical Practices
GGT	Gamma glutamyl transferase
HAS	Hunger Assessment scale
HbA1c	Hemoglobin A1c
HLGT	High Level Group of Terms
HLT	High Level Term
HR	Hazard Ratio
IC	Informed Consent
ICH	International Conference on Harmonization

ISS	Integrated Summary of Safety
ITT	Intent-to-Treat
IWRS	Interactive Web Response System
Kg	Kilogram
LDH	Lactate dehydrogenase
LLT	Lowest Level Term
LS	Least Squares
MAR	Missing at Random
max	Maximum
MedDRA	Medical Dictionary for Regulatory Authorities
MI	Multiple Imputation
mg	Milligram
min	Minimum
MNAR	Missing not at Random
NSCLC	Non-Small Cell Lung Cancer
OR	Odds Ratio
PD	Protocol Deviation
PE	Physical Examination
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PP	Per protocol
PT	Preferred Term
Q1	25th percentile or lower quartile
Q3	75th percentile or higher quartile
QD	Once daily
QC	Quality Control
RECIST	Response Evaluation Criteria in Solid Tumors
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SD	Standard Deviation
SDTM	Study Data Tabulation Model
SE	Standard Error
SI	Standard International
SMQ	Standardized MedDRA Queries
SOC	System Organ Class (in MedDRA)
TEAE	Treatment-Emergent Adverse Event
TEAESI	Treatment-Emergent Adverse Event of Special Interest
TLFs	Tables, Listings and Figures
TOI	Trial Outcome Index
WBC	White Blood Cell
WHO	World Health Organization
WHODD	World Health Organization Drug Dictionary

2. MODIFICATION HISTORY

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
Draft v0.1	10-Dec-2018	[redacted]	
Draft v0.2	29-Dec-2018		Internal review
Draft v0.3	09-Jan-2019		Incorporation of Mock tables
Draft v0.4	25-Mar-2020		Update to version3.0 of study protocol
Draft v0.5	25-Jan-2021		Update of parts related to safety and demography
Draft v0.6	02-Jun-2021		Update to version 4.0 of study protocol
Draft v0.7	17-Dec-2021		Update to version 5.A of study protocol
Draft v0.8	20-Jul-2022		Update to version 5.A/5.0 of study protocol
Draft v0.9	30-Nov-2022		Update to algorithms for missing imputation. Added new analysis for COVID-19 and exploratory analyses. Restyling of the SAP.
Final v1.0	12-May-2023		Implementation of post-BDRM updates, final rewording and formatting. Finalization of the document.

3. PURPOSE OF THE STATISTICAL ANALYSIS PLAN

This Statistical Analysis Plan (SAP) is intended to document technical and detailed specifications for the analysis of the data collected in order to generate the tables and listings to support the analysis of this phase 3 Randomized Clinical Trial, evaluating the efficacy and safety of anamorelin HCl for the treatment of malignancy associated weight loss and anorexia in adult patients with advanced Non-Small Cell Lung Cancer (NSCLC).

The structure and content of this SAP are intended to provide sufficient detail to meet the requirements identified by Regulatory Agencies and International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH): Guidance on Statistical Principles in Clinical Trials.

The SAP is also prepared in accordance with the study protocol V5.A / V5.0 and will concentrate on the detailed description of the planned analyses to be carried out to support the completion of the Clinical Study Report (CSR) for the study.

All statistical analyses will be carried out after the database is locked and treatment allocation is unblinded.

4. STUDY DESIGN

4.1. Study Design

Weight loss and anorexia (including appetite loss) represent common and debilitating manifestations in patients with advanced NSCLC. Despite the high prevalence and the association with diminished life-expectancy, the therapeutic advance in managing these symptoms has been so far limited and the treatment of malignancy associated weight loss and anorexia in cancer patients remains an area of highly unmet medical need.

The study is a multicenter, randomized, double-blind, parallel-group, placebo-controlled study to evaluate the efficacy and safety of anamorelin HCl. A total of 318 patients with advanced NSCLC with cachexia have been randomized 1:1 to anamorelin HCl 100 mg or placebo, taken orally once daily (QD) for a total of 24 weeks. Patients were instructed to take the study drug at least 1 hour before their first meal of the day. For each patient the study lasted up to approximately 27 Weeks, including the following:

- A screening period from Day -7 to -1 (on-site Visit 1)
- A 24-week treatment period (on-site Visit 2 to Visit 10)
- A 2-week follow-up period (telephone Visit 11)

A patient is defined as having completed the Study if the follow-up visit (Visit 11) was completed.

The study being placebo-controlled, all comparisons for the efficacy parameters will be done in order to demonstrate superiority of anamorelin over placebo.

Study Design and Schedule of Assessments

Assessment	Screening	Visit Day Treatment Period									Follow-up Period [2]
Day	-7 to -1	1	22 (+3)	43 (+3)	64 (+3)	85 (+3)	106 (+3)	127 (+3)	148(+3)	169 (+3)	183 (+3)
Week	-1	1	3	6	9	12	15	18	21	24	26
Visit	1	2	3	4	5	6	7	8	9	10	11
Informed consent	X										
Medical history	X										
Physical exam	X	X	X	X	X	X	X	X	X	X	
ECOG PS	X				X						
Vital signs	X	X	X	X	X	X	X	X	X	X	
Height and BMI	X										
Body weight	X	X	X	X	X	X	X	X	X	X	
12-Lead ECG	X	X	X	X	X	X	X	X	X	X	
Chemistry/Hematology	X	X	X	X	X	X	X	X	X	X	
HbA1c	X					X				X	
Urine pregnancy	X	X			X						
Urinalysis	X				X					X	
Administer FAACT questionnaire 12-item A/CS domain only	X										
Administer FACIT questionnaires[1] and Hunger Assessment Scale questionnaires		X	X	X	X	X	X	X	X	X	
Administer PGIS questionnaires for body weight and anorexia	X			X	X						
Administer PGIC questionnaires for body weight, anorexia and overall condition,				X	X						
CT scan	X	→	→	→	→	X	→	→	→	→	
Adverse events	X	X	X	X	X	X	X	X	X	X	X
Prior and concomitant medications (including NSCLC treatment)	X	X	X	X	X	X	X	X	X	X	
Randomization		X									
Dispense new and/or return blister card for study drug dosing; check compliance		X	X	X	X	X	X	X	X	End	

1) FACIT questionnaires: includes all 4 domains, PWB, FWB, SWB, and EWB of FACT-G, 12-item A/CS domain and 13-item FACIT-F fatigue domain

2) Follow-up period: by telephone

4.2. Treatments

Group 1 / Test group – 100 mg anamorelin HCl (once per day)

Group 2 / Control group – placebo (once per day)

4.3. Questionnaires

The derivations of all the scores and algorithms described in the present section will be obtained directly by the eCOA/ePRO vendor (Clario, formerly eResearch Technology). No further calculation or derivation will be done at SDTM or ADaM level. The derived variables will be reported in SDTM.QS and ADaM.ADQOL domains after confirmatory crosschecks performed by the CRO and Helsinn.

Algorithm will be reported in SDTM Define.

4.3.1. Functional Assessment of Anorexia/Cachexia Therapy (FAACT)

The FAACT measure is a 39-item questionnaire which combines the 27 items of the FACT-G (Functional Assessment of Cancer Therapy-General) questionnaire to the 12-item measure of the FAACT A/CS which examines patients' perception of Anorexia Cachexia symptoms and concerns.

Each item is scored by the patient on a 5-point Likert-type scale, ranging from 0 (Not at All) to 4 (Very Much).

Details on scoring are reported in Appendix 12.3.

4.3.2. Functional Assessment of Chonic Illness Therapy – Fatigue (FACIT-F)

The FACIT-Fatigue subscale is a 13-item measure which is summed to the 27-item FACT-G score to calculate the FACIT-F Total score of the 40-item FACIT-F measure; each item of the FACIT-Fatigue scale is to be scored by the patient on a 5-point Likert-type scale, ranging from 0 (Not at All) to 4 (Very Much).

Details on scoring are reported in Appendix 12.3.

5. OBJECTIVES

The primary objective of the present ANAM-17-20 study is to demonstrate superiority of anamorelin vs. placebo on the gain in body weight and improvement of anorexia symptoms in patients with advanced non-small cell lung cancer (NSCLC) with BMI<20 kg/m2 and involuntary weight loss of more than 2% in the 6 months preceding screening. The secondary objective is to evaluate the safety and tolerability of anamorelin HCl, and to further evaluate anamorelin efficacy profile.

5.1. Estimands

The analysis to address this objective has been developed in consideration of ICHE9(R1). Specifically, two intercurrent events have been identified: study drug discontinuation and death.

As to the study drug discontinuation, the “treatment policy” strategy will be applied to all efficacy endpoints (primary/secondary/exploratory) declared in Section 6 (apart from HAS and PGIS/PGIC). Efforts have been made to retain all patients in study follow-up, whether or not they have discontinued the study treatment. In view of the statistical analysis, methods for missing imputation will be applied in case of assessments not performed after study discontinuation (Appendix 12.1).

In the target population of patients with advanced NSCLC it is expected that death will be mostly due to the progression of the underlying malignant disease. The aim of anamorelin treatment is to increase body weight and improve anorexia symptoms while the patient is alive. Hence, it is necessary and sufficient to obtain regular assessments on body weight and anorexia symptoms between randomization and death. Therefore, a “while-on-treatment” type strategy for death (i.e. while alive) is applied.

6. STUDY ENDPOINTS

6.1. Efficacy Endpoints

6.1.1. Primary Efficacy Endpoints

The co-primary efficacy endpoints are:

- Mean change in body weight from baseline over 12 weeks.
- Mean change in 5-item Anorexia Symptom Subscale (5-IASS) from baseline over 12 weeks.

6.1.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

- Duration of treatment benefit in weight gain (≥ 0 kg) over 12 weeks.
- Duration of treatment benefit in weight gain (≥ 1.5 kg) over 12 weeks.
- Duration of treatment benefit in anorexia symptoms (≥0 points) over 12 weeks, as measured by the 5-item Anorexia Symptom Subscale.
- Duration of treatment benefit in 5-item Anorexia Symptom Subscale (increase ≥ 3 points) over 12 weeks.
- Mean change in FAACT 12-item A/CS domain from baseline over 12 weeks.
- Mean change in FACIT-F 13-item subscore from baseline over 12 weeks
- Mean change in FAACT total score from baseline over 12 weeks.

6.1.3. Exploratory Efficacy Endpoints

- Mean change in body weight from baseline over 6, 9, 15, 18, 21, and 24 weeks.
- Mean change in patient-reported anorexia symptoms from baseline over 6, 9, 15, 18, 21, and 24 weeks, as measured by the 5-item Anorexia Symptom Subscale.
- Mean change in FAACT total score from baseline over 6, 9, 15, 18, 21, and 24 weeks.
- Composite Clinical Response (CCR) at Week 3, 6, 9, 12, 15, 18, 21, and 24.
- Change in body weight from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24.
- Change in patient-reported anorexia symptoms from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24, as measured by the 5-item Anorexia Symptom Subscale.
- Clinically meaningful gain in body weight (1.5 kg) from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24.
- Clinically meaningful increase in 5-item Anorexia Symptom Subscale (3 points) from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24.

- Percentage of change in body weight from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24.
- Changes from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24 in the FAACT total score, FACT-G total score, FAACT TOI, the 12-item A/CS domain and the 4-item Anorexia Concerns Subscale.
- Changes from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24 in the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) domain as well as the FACIT-F TOI score.
- Changes from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24 in the Hunger Assessment Scale.
- Changes from baseline to Week 6 and 9 in the Patients' Global Impression of Severity (PGIS) scale and the Patients' Global Impression of Change (PGIC) scale at Week 6 and 9.
- Changes from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24 in the following items extracted from Physical Well-Being and Functional Well-Being domains:
    - GP1, GP7, GF5, GF6, GF7.
- Changes from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24 in the following items extracted from the 12-item A/CS domain:
    - ACT1, ACT2, ACT4, ACT13.
- Changes from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24 in the following items extracted from the FACIT-F (Fatigue) questionnaire:
    - HI7, HI12, AN2, AN5, AN12.
- Correlation between Weight gain and improvement on the patient reported anorexia score from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24.
- Correlation between Weight gain and improvement on the patient reported anorexia score from baseline over 12 weeks.

6.2. Safety Endpoints

- Treatment-emergent adverse events (TEAEs) and serious adverse events (SAEs)
- Changes from Baseline in vital signs, ECG parameters and laboratory assessments (chemistry, hematology, and urinalysis)
- Shifts from Baseline in laboratory assessments, physical examination and ECG overall interpretation.
- Overall Survival
- Overall Response based on tumor assessment

7. STUDY METHODS AND STATISTICAL PRINCIPLES

7.1. Randomization

Central randomization (IWRS) was stratified by:

- Line of systemic anti-cancer treatment (i. first vs ii. second vs iii. third line or higher),
- Type of anti-cancer therapy (immunotherapy vs non-immunotherapy),
- Baseline score of 5-item Anorexia Symptom Subscale (≤10 vs >10).

7.2. Sample Size

At the time when the trial was initiated, the primary endpoint was the Composite Clinical Response (CCR) rate. The sample size was determined based on results obtained in patients participating to the previous trials (HT-ANAM-301 and HT-ANAM-302 (4), (5)). The retrospectively observed CCR rates, in the subset of patient with baseline BMI <20 kg/m² and ongoing problems with appetite/eating as defined by inclusion criterion 5, were 20% for anamorelin and 7.5% for placebo at week 12, and 21.0% vs. 6.0% at week 9.

To be conservative the study was adequately powered to detect an improvement in CCR Rate at week 9 from 7.5% in the placebo group to 20.0% for the anamorelin group. This improvement would be detected with 90% power, using a logistic regression for the CCR, with a sample size of 316 patients (158 patients per arm) and a 2-sided 0.05 significance level.

In the latest version (V5.0-V5.A) of the study protocol, co-primary endpoints were defined as follows:

- mean change in body weight over 12 weeks
- mean change in 5-IASS over 12 weeks.

With the fixed sample size of 158 patients per arm, study power was computed by means of simulations based on patients participating to previous company trials comparing anamorelin and placebo. Data from these trials were resampled in an unbalanced and in a balanced way with respect to 5-IASS at baseline (<= or >10), as those studies were not stratified and were unbalanced between treatment groups in terms of 5-IASS. To explore the power of the test procedure for the new endpoint, many scenarios to set assumptions for missing data were simulated, based on choices between different options for the following topics:

- The patient's baseline characteristics of the sample to be obtained (with BMI≤20, ACS-12≤37, 5-IASS≤17, fulfilling inclusion criteria 4 and 5, with/without missing data at baseline, with/without 5-IASS missing at week 3)
- The kind of imputation method adopted (fully MAR or MNAR+MAR based on predefined rules)

- The variables to be included as covariates in the imputations (from a large number to no covariates)
- The process for the MNAR+MAR imputation method: the different sets of subjects contributing to the imputations (all subjects with data vs. all complete subjects) and different sequential steps for applying MAR and MNAR
- The method of analysis (ANOVA with the 5-IASS stratum included in the ANOVA model, then including/excluding in the model the baseline covariate of the variable under analysis)

For each scenario the number of simulated studies ranged from 400 to 1000.

The simulations showed a power close to 100% to test an effect size in terms of mean change in body weight over 12 weeks approximately equal to 1.8 kg. The power to test an effect size in terms of mean change in 5-IASS over 12 weeks within the range of 1.2 – 1.4 was showed to range between 78% and 87%. This leads to a joint power for the two co-primary endpoints of at least 78%.

7.3. Statistical Testing

The primary and the secondary efficacy endpoints will be evaluated using a hierarchy for testing in order to control type I error rate. The order of testing will be as follows:

- The mean change in body weight over 12 weeks ($H0_W$) and the mean change in 5-item Anorexia Symptom Subscale over 12 weeks ($H0_A$), assessed as co-primary endpoints; Definition and derivation of the co-primary efficacy endpoints is given in Section 10.1.1.
- The secondary efficacy endpoint related to the duration of treatment benefit over 12 weeks in body weight gain (≥ 0 kg) ($H0_{d0W}$).
- The secondary efficacy endpoint related to the duration of treatment benefit over 12 weeks in body weight gain (≥ 1.5 kg) ($H0_{dtW}$).
- The secondary efficacy endpoint related to the duration of treatment benefit over 12 weeks in 5-item Anorexia Symptom Subscale (≥0 points) ($H0_{d0A}$).
- The secondary efficacy endpoint related to the duration of treatment benefit over 12 weeks in 5-item Anorexia Symptom Subscale (≥ 3 points) ($H0_{dtA}$).
- The secondary efficacy endpoint related to the mean change in FAACT 12-item A/CS domain score from baseline over 12 weeks ($H0_{mF}$).
- The secondary efficacy endpoint related to the mean change in FACIT-F ($H0_{mFF}$) from baseline over 12 weeks.
- The secondary efficacy endpoint related to the mean change in FAACT total score ($H0_{mFt}$) from baseline over 12 weeks.

Definition and derivation of the secondary efficacy endpoints is given in Section 10.2.1.

Each of these tests will be done sequentially at the level of 0.05 bilateral.

7.3.1. Testing the co-primary endpoints

To preserve the experimental Type I error rate (threshold of 0.05 two-sided), in order to declare the test component (anamorelin) superior to the placebo, both co-primary endpoints must be significant, meaning that both $H0_w$ and $H0_a$ must be rejected simultaneously.

- The null hypothesis for the first co-primary endpoint of mean change in body weight from baseline over 12 weeks (H_{0w}) and the corresponding alternative hypothesis (H_{1w}) are defined as:

  H_{0w}: $MW_a = MW_p$

  H_{1w}: $MW_a \neq MW_p$

Where MW_a is the mean change in body weight from baseline over 12 weeks for the anamorelin arm and MW_p is the mean change in body weight from baseline over 12 weeks for the placebo arm.

- The null hypothesis for the other co-primary endpoint of mean change from baseline over 12 weeks in patient 5-item Anorexia Symptom Subscale (H_{0A}) and the corresponding alternative (H_{1A}) are defined as:

  H_{0A}: $MA_a = MA_p$

  H_{1A}: $MA_a \neq MA_p$

Where MA_a is the mean change from baseline over 12 weeks in 5-item Anorexia Symptom Subscale for the anamorelin arm and MA_p is the mean change from baseline over 12 weeks in 5-item Anorexia Symptom Subscale for the placebo arm.

7.3.2. Testing the secondary endpoints

- The null hypothesis concerning the first secondary endpoint of duration of treatment benefit over 12 weeks in body weight (≥0 kg) ((H_{0d0w}) and the corresponding alternative hypothesis (H_{1d0w}) are defined as:

  H_{0d0w}: $D_0W_a = D_0W_p$

  H_{1d0w}: $D_0W_a \neq D_0W_p$

Where D_0W_a is the duration of treatment benefit over 12 weeks in body weight gain (≥ 0 kg) for the anamorelin arm and D_0W_p is the duration of treatment benefit over 12 weeks in body weight gain for the placebo arm.

H_{0d0W} will be tested using a two-sided Type-I error threshold of 0.05 only if both H_{0w} and H_{0a} (i.e., the null hypothesis concerning the primary endpoints) are rejected.

- The null hypothesis concerning the secondary endpoint of duration of treatment benefit in body weight over 12 weeks (≥ 1.5 kg) (H_{0dtW}) and the corresponding alternative (H_{1dtW}) are defined as:

  H_{0dtW}: $D_tW_a = D_tW_p$

  H_{1dtW}: $D_tW_a \neq D_tW_p$

Where D_tW_a is the duration of treatment benefit over 12 weeks in weight (≥ 1.5 kg) for the anamorelin arm and D_tW_p is the duration of treatment benefit over 12 weeks in weight for the placebo arm.

H_{0dtW} will be tested using a Type-I error threshold of 0.05 two sided only if H_{0d0W} (i.e., the null hypothesis concerning the first secondary endpoint) is rejected; note the latter rejection requiring the rejection of if both H_{0w} and H_{0a}.

- The following hypothesis systems will then be tested based on the rules above described.

  H_{0d0A}: $D_0Aa = D_0A_p$

  H_{1d0A}: $D_0A_a \neq D_0A_p$

Where D_0A_a is the duration of treatment benefit over 12 weeks in 5-item Anorexia Symptom Subscale (≥ 0 points) for the anamorelin arm and D_0A_p is the duration of benefit over 12 weeks in 5-item Anorexia Symptom Subscale for the placebo arm.

H_{0dtA}: $D_tAa = D_tA_p$

H_{1dtA}: $D_tA_a \neq D_tA_p$

Where D_tA_a is the duration of treatment benefit over 12 weeks in 5-item Anorexia Symptom Subscale (≥ 3 points) for the anamorelin arm and D_tA_p is the duration of benefit over 12 weeks in 5-item Anorexia Symptom Subscale for the placebo arm.

H_{0mf}: $MF_a = MF_p$

H_{1mf}: $MF_a \neq MF_p$

Where MF_a is the mean change from baseline over 12 weeks in FAACT 12-item A/CS domain score in the anamorelin arm and MF_p is the mean change from baseline over 12 weeks in FAACT 12-item A/CS domain score in the placebo arm.

Similar hypothesis systems are set-up for FACIT-F (H_{0mFF}, H_{1mFF}) and FAACT total score (H_{0mFt}, H_{1mFt}).

8. GENERAL CONSIDERATION FOR STATISTICAL ANALYSIS

8.1. Descriptive Statistics

All analyses will be conducted using SAS version 9.4 or higher.

The following conventions will be used when presenting summary statistics:

- Continuous variables will be summarized using the following descriptive statistics: number of subjects (n), mean, standard deviation (SD) median, minimum (min), maximum (max), 25th percentile (lower quartile or Q1) and 75th percentile (upper quartile or Q3).
- Categorical and binary variables will be summarized by the number of subjects (n) and the percent of subjects in each category. For AE tables, the number of events (E) will also be presented. Percentages will be presented to 1 decimal place.
- The number of subjects or observations will be presented as an integer.
- Mean, lower quartile and upper quartile will be presented to 1 decimal place more than the raw value.
- Min, max and median will be presented to the same number of decimal places as the raw value.
- Standard deviation will be presented to 1 decimal place more than the mean value, up to a reasonable number of decimal places (i.e., rarely more than 3 decimal places unless the nature of the data suggests this).
- The denominator for percentage calculations will be the number of non-missing observations at that visit. Where this isn't appropriate, the number of subjects in the relevant analysis set will be used.

8.2. Addressing intercurrent events

8.2.1. Intercurrent event death

The data from the previous trials HT-ANAM-301 (4) and HT-ANAM-302 (5) suggest a positive effect of anamorelin on the weight and the 5-IASS. Death from underlying disease is rather inevitable in this target population and it is anticipated to not reflect a failure of study treatment. In accordance with the ICHE9 R1 guideline, the death is considered as an intercurrent event after which no value is measurable. In order to measure the treatment effect for these patients, the "while alive" strategy will be used. Therefore, for patients dying before the completion of the period, no imputation strategy will be adopted for outcomes after death and only values collected before death will be used for the evaluation of the endpoint.

8.2.2. Intercurrent event study drug discontinuation

The treatment policy strategy will apply to address study drug discontinuation.

Due to the nature of the patients' underlying disease, a non-ignorable amount of missing data is expected in this trial. Every effort has been made to collect data from the protocol-specified assessments; patients were asked to continue to participate in the study and complete planned visits up to the end of the study period (Week 24) even if they have permanently discontinued treatment with the IMP ("retrieved dropout" approach).

Reason for treatment discontinuation will be collected in the CRF and will be considered for use in investigating possible unexpected patterns of missing data. To minimize the risk of bias in the analysis, specific rules for single/multiple imputation will be adopted

8.3. Handling of missing data

8.3.1. Efficacy Missing Data

The preferred statistical approach to manage the missing values due to the intercurrent event was the Multiple Imputation (MI) using retrieve dropouts (e.g.: assessments in patients no more under randomized treatment). However, at the BDRM it was confirmed that only a small number of retrieved dropouts was available, making unfeasible the established approach. For this reason, a control-based MI approach will be adopted by considering a Missing Not at Random (MNAR) mechanism for missing data: imputation of values in the active arm will be done using the non-missing values from the control group. This approach does not assume benefits for anamorelin in case of intercurrent event and limits a post-discontinuation clinical effect to that of placebo.

Efficacy missing data will be imputed on the bases of rules described in Appendix 12.1.

For missing data the imputation of the values will be done at the level of the visit for the raw values (for instance weight).

Post-discontinuation missing date of visit X will be imputed by adding 3 weeks to the date of the previous available visit X-1. The date of visit X+1 will be imputed by adding 3 weeks to the date of Visit X imputed in the previous step and so on.

Imputed values will then be used to derive primary, secondary and exploratory endpoints according to rules reported in Section 10.1.1, 10.2.1 and 10.3.1.

Datasets including imputed values and resulting derived variables will be used for all efficacy endpoints (primary, secondary and exploratory).

8.3.2. Missing Date on concomitant medications

In the determination of whether a medication is taken during the Treatment Period, the following rules for missing and partial dates will be applied:

- If the end date is before the first dose of study medication, then the medication is a prior medication, regardless of whether or not the start date is missing or partially missing

- If the end date is on or after the date of the first dose of study medication, then the following rules will be applied for missing or partially missing start dates:
  - If the start date is completely missing, then the medication will be assumed to be concomitant.
  - If only the year of the start date is recorded and the year is either before or the same year as the date of last dose of study medication then the medication will be assumed to be concomitant.
  - If only the year of the start date is recorded and the year is after the year of the date of last dose of study medication, it will be considered as a post-study treatment medication.
  - If only the year and the month of the start date is recorded and the month specified is either before or the same month as the date of last dose of study medication, then the medication will be assumed to be concomitant.
  - If only the year and the month of the start date is recorded and the month specified is after the month of the date of last dose of study medication, it will be considered as a post-study treatment medication.

The same rules are applicable for the concomitant procedures.

8.3.3. Questionnaires Missing Data

See Section 4.3 for the questionnaires and Appendix 12.3 for Scoring guidelines.

In cases where individual items are skipped, subscale scores can be prorated using the average of the other answers in the scale. A FACIT measure is considered to be an acceptable indicator of patient quality of life as long as overall response rate at item level is greater than 80% (e.g., at least 22 of 27 FACT-G items completed). This is not to be confused with individual subscale item response rate, which allows a subscale score to be prorated for missing items if greater than 50% of items are answered e.g., a minimum of 4 of 7 items, 4 of 6 items, etc. The total score is then calculated as the sum of the un-weighted subscale scores. In addition, a total score should be calculated only if ALL of the component subscales have valid scores.

PGIC and PGIS questionnaires have no score derivation.

8.3.4. Overlapping/Duplicated Diaries

In some patients questionnaires recorded at screening visit (expected to be done between Day -7 and Day -1) were actually filled in on Day 1, just before the questionnaire filled in for the Visit 2, meaning that the date of screening assessment is the same as the date of Visit 2. These cases were checked during BDRM and minor protocol deviations were recorded.

In some patients the overlapping of assessment dates happened between Visit 3 and Visit 10. The same questionnaire was performed twice (with a lag of few minutes) in a day (usually corresponding to the Day of a given Visit X). The second questionnaire

was entered as performed at the Visit X, while the first one was entered as referring to the prior Visit (X-1). For these cases the following rule will be applied:

If the date of assessment [Visit (X-1)] >= date of assessment Visit (X),

then the date and collected value of questionnaire assigned to Visit X-1 will be set to missing. The date of assessment Visit (X-1) will be subsequently reset to the date of Visit (X-1) from Study Visits domain; this will be done at ADaM level. Questionnaire value set to missing will be imputed according to the rules described in Appendix 12.1.

If a mismatch between the date of score assessment (i.e.: questionnaire date aforementioned) and the date of corresponding individual items assessment is detected, the date of score will be replaced with the date of individual items. If a mismatch between the dates of individual items assessment is detected, the maximal date will be used to replace the date of score.

8.3.5. Missing Date on Adverse Events

In the determination of whether an Adverse Event is Treatment Emergent (TEAE), the following rules for missing and partial dates will be applied:

- If an incomplete onset date was collected for an AE, the event will be considered TEAE unless there is other evidence that confirms that the event was not TEAE (e.g.; next bullet point)
- If the AE end date is before the first dose of study medication, then the AE is pre-treatment, regardless of whether or not the AE start date is missing or partially missing.

8.4. Covariates

The three stratification factors at randomization are used as categorical covariates in each statistical model:

- Systemic anti-cancer treatment line (first line vs second line vs third or higher line)
- Cancer treatment (immunotherapy vs no immunotherapy)
- The 5-item Anorexia Symptom Subscale (≤10 vs >10).

8.5. Interim Analysis

Not applicable. No formal interim analysis is planned for this study.

8.6. Subgroups analyses

Subgroup analyses (specifically the analysis of the coprimary and secondary efficacy endpoints) will be performed for the subgroups for which numerical feasibility was confirmed at the BDRM, as defined by the following variables:

- Age (≤ 65 years or > 65 years)
- Gender
- Systemic anti-cancer treatment line (first line vs second line vs third or higher line) – stratification factor
- Cancer treatment (immunotherapy vs no immunotherapy) – stratification factor
- The 5-item Anorexia Symptom Subscale (≤10 vs >10) – stratification factor
- Patients who completed 24 weeks treatment

8.7. Visit windows

For the endpoints based on mean change and duration of benefit, the actual day of assessment will be used for the computation. No other visit windowing is applicable and planned to be done apart from corrections related to the overlapping/duplicated diaries described in Section 8.3.4.

8.8. Definitions

Baseline

The baseline value for all efficacy analyses is defined as the value collected at Visit 2 (randomization); if this is not available or collected day(s) later as compared to Visit 2, the value collected at screening visit will be used.

For the following assessments, baseline is the value reported at Screening (V1):

- ECOG PS
- HbA1c
- all Urinalysis parameters
- the Patients' Global Impression of Severity (PGIS) scale
- the Patients' Global Impression of Change (PGIC) scale

The baseline value for all safety analyses is defined as the last measurement performed before study drug administration.

Change from Baseline

Any "Change from baseline" variable will be computed as:
"variable" at Visit X – Baseline,
with X≥3 and Baseline defined as above.

Last assessment

Last assessment performed (body weight / 5-IASS) regardless of treatment discontinuation.

Retrieved dropout

Patient who drops the treatment but who continues to be assessed.

Non-monotone pattern of missingness (also called "**Intermittent missing values")**

Missing values may occur in the middle of a subject record (e.g., missed intermediate visit). It is expected that a value before the missing assessment, and a value after the missing assessment are available.

Monotone pattern of missingness (also called "**non-Intermittent missing values**") Missing values are located in a block at the end of data records (e.g., due to study discontinuation).

Study Treatment Start Date
It is the date of the first intake of the study treatment.

End of Treatment (EOT) Date
It is the date declared in the End of Treatment CRF Form. This date will be used in the derivations unless specific adjustments are needed (e.g., for actual exposure derivation).

Last follow-up contact date
It is the date of the last contact for patients alive. It is the date of last follow-up in case of final phone contact (Visit 11 for patients completing the study, before Visit 11 in case of treatment discontinuation), or the date of last performed visit in case of patient lost to follow-up.
It corresponds to the date of death in case a patient died.

Study completion
A patient will be defined as having "completed" the study when the patient completes Visit 11. Termination of the study before performing Visit 11 will be considered as a premature study discontinuation.

Patients who completed 24 weeks treatment
A patient will be defined as having "completed" 24 weeks treatment when the patient received study drug at the day of Visit 10 (either on-site or at home) and have evidence of treatment for all previous weeks.

Fluid retention
Presence of fluid retention will be assumed when either a treatment emergent AE associated with fluid retention is present or edema is recorded in the CRF by investigator at time of the assessment of weight. Specific assessments are scheduled at each study visit.

After a specific review of the LLTs during the BDRM, AEs coded with the following PTs were confirmed as possibly impacting the weight measurements: 'Fluid retention', 'Ascites', 'Pleural effusion', 'Oedema peripheral', 'Oedema'. These AEs will be taken into account.

Clinically meaningful thresholds for 5-IASS change and Body Weight Change
To support with patient informed interpretation of meaningful change or stabilization, the results of a quantitative analysis of a separate study (7), combined with evidence from the embedded qualitative interview sub-study (6), support a responder definition of 3.0 for the 5-IASS Score and 1.5 kg weight gain for body weight.

These thresholds were defined in line with FDA guidance, therefore anchor-based methods for score interpretation were considered the primary analysis. Distribution-based methods were used as supportive and secondary analysis using PGIS (Patient Global Impression of Severity) and PGIC (Patient Global Impression of Change) as anchors.

For the analysis of the duration of treatment benefit the definition of a threshold to establish a change from baseline as clinically meaningful is provided (Sections 10.2 and 10.2.1)

Last value
For Laboratory Shift Tables analysis, the Last Value is the last value reported by the patient during the study period.

8.9. Other Considerations: Handling impact of COVID-19 pandemic

The present document and the analysis described in it, have been updated in compliance with the FDA **guidance** contained in the Conduct of Clinical Trials of Medical Products during the Covid-19 Public Health Emergency (Version March 2020, updated on August 30th 2021).

The Sponsor provided instructions to sites during the study in order to manage and mitigate the COVID-19 pandemic effects on the study, particularly allowing remote subject visits (refer to the Guidance and contingency plan from Sponsor to Sites (8)).

8.9.1. Handling Impact of COVID-19 in the Statistical Analysis

All deviations related to the COVID-19 pandemic and resulting from the measures described in the Sponsor's contingency plan have been captured and discussed during the BDRM.

Sensitivity analyses have been planned based on the following considerations. Upfront it can be stated that:

- A tangible impact on measurements of body weight can be expected. Body weight values were reported by patient at each visit during a phone call. Patients were asked to use their own weight scale (analog or digital). This implies a lack of standardization and therefore to a certain degree of variability and lower accuracy.
- A minor impact is anticipated on eCOA endpoints. During remote visit on the phone, patient was asked to respond to each item of each questionnaire as per schedule at the given visit by delegated study personnel who entered responses in eCOA system accordingly, and data underwent to standard transmission process.
- There is no reason to expect any worsening in patient safety.

9. STUDY POPULATION

9.1. Trial Subjects

9.1.1. Analysis Sets

The following study populations (analysis sets) will be considered for analysis:

- The Intent-to-Treat (ITT) Set includes All randomized patients and will be analyzed as per planned treatment.
- The Full Analysis Set (FAS) includes all randomized patients who took at least one dose of trial medication and for whom any post-randomization data (in addition to the trial medication intake) are available. The FAS will be analyzed as per randomization.
- The Safety Set includes all patients who receive any study drug and will be analyzed as per actual treatment.
- The Per-Protocol (PP) set is subset of the randomized patients who do not have major protocol violations and will be analyzed as per actual treatment.

9.1.2. Protocol Deviations

Protocol deviations (PDs) were collected and documented by the study monitors and medical team throughout the study period. During the data cleaning process, PDs detectable programmatically in the clinical database were also identified and documented for final discussion and categorization during the BDRM.

Exact definition of major and minor protocol deviations affecting primary efficacy were specifically discussed and approved by Helsinn at the BDRM and documented in the meeting minutes.

The finalization of protocol deviations and exclusions from the Per Protocol set have been made prior to randomization code being revealed.

Minor and major protocol deviations identified will be summarized by treatment group and for all patients in the ITT Set

A listing of all protocol deviations will also be provided.

9.1.3. Disposition of Subjects and Discontinuations

All summaries will be presented on the ITT Set.

The number of subjects treated in the study will be presented, together with the number and percentage of subjects who completed the study and the number and percentage of subjects who withdrew from the study prematurely; a breakdown of the corresponding reasons for withdrawal will be presented. The number of subjects who discontinued the study treatment prematurely will also be presented with a breakdown of the corresponding reasons for study treatment discontinuation.

A tabulation of the number and percentage of subjects randomized at each site and country will be summarized for all patients in the ITT Set.

9.2. Demographics and Other Baseline Characteristics

9.2.1. Demographics

Subject demographics will be summarized using descriptive statistics for continuous variables and frequency distributions (number and percentage of subjects) for categorical variables. Summaries will be provided overall and by treatment group at Screening for all defined analysis sets.

The following demographic data and baseline characteristics will be summarized:

- Age at Screening
- Gender
- Race and Ethnicity
- Height
- Weight at Screening
- BMI at Screening

If subjects have record of more than one race, then for the summary table, these subjects will be included in the Other Race group. BMI will be calculated from the height and weight values.

All demographic data and baseline characteristics will be listed for all patients in the ITT Set. In the demography listing, if a subject has more than one race recorded, then all races selected will be listed for that subject.

9.2.2. ECOG Status at screening

ECOG status at baseline (screening) will be summarized using frequency distributions (number and percentage of subjects). Summaries will be provided overall and by treatment group for all defined analysis sets.

ECOG status will also be listed for all patients in the ITT Set.

9.2.3. Stratification factors at V2 (randomization)

Chemotherapy line, Anti-cancer treatment (Immunotherapy/Non-immunotherapy) and 5-IASS score used as stratification factors at randomization will be summarized using frequency distributions (number and percentage of subjects) for categorical variables. Summaries will be provided overall and by treatment group for ITT set only.

All stratification factors will also be listed for all patients in the ITT Set.

9.2.4. Body Weight and 5-IASS at V1 (Screening) and V2 (Randomization)

Body Weight and 5-IASS score used as co-primary endpoint for the efficacy analysis, will be summarized using descriptive statistics for continuous variables. Summaries for screening and randomization visits will be provided in a single table overall and by treatment group for ITT set.

Body Weight and 5-IASS score at V1 and V2 will also be listed for all patients in the ITT Set.

9.2.5. Vital Signs at screening

Vital Signs collected at screening will be summarized using descriptive statistics for continuous variables. Summaries will be provided overall and by treatment group for all defined analysis sets.

The following Vital Signs data will be summarized:

- Systolic Blood Pressure
- Diastolic Blood Pressure
- Pulse Rate
- Respiratory Rate
- Body Temperature

All data will be listed for all patients in the ITT Set.

9.2.6. Reproductive Status

Reproductive status information will be summarized using frequency distributions (number and percentage of subjects) for categorical variables. Summaries will be provided overall and by treatment group at Screening for ITT and Safety sets.

The following reproductive status information will be summarized:

- Patient childbearing potential
- Use of reliable contraceptive measures
- Reason of not being childbearing potential

All data will be listed for all patients in the ITT Set.

9.2.7. Prior Body Weight Loss at screening

Body Weight loss within 6 months prior to screening will be summarized using descriptive statistics for continuous variables. Summaries will be provided overall and by treatment group at Screening for all defined analysis sets.

All data will also be listed for all patients in the ITT Set.

9.2.8. Medical History and Concomitant Diseases

Medical history and concomitant diseases will be coded with MedDRA dictionary (version 22.0).

Medical history refers to all conditions that started and ended before the first dose of study drug.

Concomitant diseases refer to all conditions started before the first dose of study drug and with an end date on or after the first dose of study drug.

Medical history and concomitant diseases will be summarized in frequency tables by SOC and PT (sorted by descending frequency of SOC and then PT in the anamorelin arm) for ITT and Safety Sets.

A listing of Medical History and Concomitant diseases will also be provided for the ITT Set.

9.2.9. NSCLC status at Diagnosis and study Entry

NSCLC status at Diagnosis will be summarized using descriptive statistics for continuous variables and frequency distributions (number and percentage of subjects) for categorical variables. Summaries will be provided overall and by treatment group at Screening for ITT and Safety sets.

The following information will be summarized:

- Time to initial diagnosis at study screening (Year)
- Stage at Diagnosis (TNM classification)
- Overall stage
- Site of Metastasis
- Mutational Status and type of mutation
- TUMOR PD-L1 Expression

The same descriptive statistics will be provided for NSCLC status at Study Entry (excluding Time to initial diagnosis at study screening).

All NSCLC status data at diagnosis and study entry will be listed for the ITT Set.

9.2.10. Physical examination

Physical examination data will be summarized using frequency distributions (number and percentage of subjects). Each Body System examined (General Appearance, Head, Eyes, Ears, Nose and Throat, Skin, Neck and so on) will be provided overall and by treatment group at Screening for ITT and Safety sets.

All physical examinations will also be listed for all patients in the ITT Set.

9.3. Prior and Concomitant Medications/Procedures

Medication usage will be coded using the World Health Organization (WHO) Drug Dictionary (version WHO Drug Enhanced version Mar 2019 C3). Medications will be presented by WHO Drug Anatomical/Therapeutic/Chemical category and Active ingredient.

Summaries will be presented for prior and concomitant medications and procedures. All summaries will present the number and percentage of subjects for each medication/procedure.

Prior and concomitant medication/procedures will be described as first excluding NSCLC treatment/procedures, then only the cancer treatment (e.g.: chemotherapy) or procedures (e.g.: radiotherapy or surgery) will be presented.

The denominators for calculating the percentages will be based on the number of subjects in the ITT set.

Summaries will be provided overall and by treatment group for ITT and Safety sets.

All prior and concomitant medications/procedures presented in this section will be listed for patients in the ITT Set.

9.3.1. Prior/Concomitant Medications (excluding NSCLC treatment)

Medications (excluding the NSCLC treatment) taken with an end date occurring before the first dose of study drug will be identified as prior medications.

Medications (excluding the NSCLC treatment) taken with a start date occurring on or after the first dose date or medications taken with a start date prior to the first dose date and an end date on or after the first dose date and a start date before EOT date will be identified as concomitant medications.

9.3.2. Prior/concomitant Procedures (excluding NSCLC treatment)

Procedures (excluding the NSCLC treatment) performed with an end date occurring before the first dose of study drug will be identified as prior procedures.

Procedures (excluding the NSCLC treatment) performed with a start date occurring on or after the first dose date or procedure performed with a start date prior to the first dose date and an end date on or after the first dose date and a start date before EOT date will be identified as concomitant procedures.

9.3.3. Prior/Concomitant Cancer treatment

Cancer treatment medications taken with an end date occurring before the first dose of study drug will be identified as prior cancer treatment.

Cancer treatment medications taken with a start date occurring on or after the first dose date or medications taken with a start date prior to the first dose date and an end date on or after the first dose date and a start date before EOT date will be identified as concomitant cancer treatment.

9.3.4. Prior/concomitant Radiotherapy/Surgery

Procedures (considering only those for cancer treatment, e.g: radiotherapy/surgery) performed with an end date occurring before the first dose of study drug will be identified as prior procedures.

Procedures (considering only those for cancer treatment, e.g: radiotherapy/surgery) performed with a start date occurring on or after the first dose date or procedure performed with a start date prior to the first dose date and an end date on or after the first dose date and a start date before EOT date will be identified as concomitant procedures.

9.4. Treatment Compliance

Compliance will be evaluated overall for the Safety Set, ITT Set and Per Protocol Set.

Treatment compliance will be calculated as:

Compliance (%) = (# tablets taken / # tables to be taken according to the study schedule).

Where:

tables taken = total number of tables taken between the date of first intake (included) and the date of last intake (included), as collected in the drug accountability CRF form.

tables to be taken according to the study schedule = date of Visit 10 - date of Visit 2 +1.

Date of Visit 10 will be replaced by the date of premature treatment discontinuation for patients who discontinued treatment prematurely.

In case a patient completed the treatment but missed Visit 10 or in case a patient prematurely discontinued treatment and presented an EOT date, as recorded in the related CRF form, which is considered as implausible (i.e.: EOT date greater than 25 days after the date of the first dose of the last kit), then the corresponding Visit 10 date or EOT date will be replaced with the date of last visit performed + 24 days.

If a subject returns all study medication that was dispensed, then the subject will not be classified as treated.

In case of missing drug accountability data, it will be assumed that no tablets were taken for the corresponding days with missing data.

For patients with kit(s) not returned, the information available in the drug administration CRF form about the documented tablet intake during the applicable study visit (i.e.: the study visit during which the first tablet from the kit – then never returned – was administered) will be used to add such one single day of documented treatment. For patients with kit(s) not returned and not having any information in the drug administration CRF form, i.e., there is no evidence that at least one tablet was taken, the number of tables taken will be assumed as 0.

A patient will be considered compliant with the therapy if his/her overall compliance is higher or equal to 80%.

Compliance will be reported numerically using descriptive statistics and categorically (using categories <80%, ≥80% to ≤100%, and >100%), by treatment group.

10. EFFICACY ANALYSIS

All efficacy analyses will be performed on the ITT population.

Some supportive analyses of the primary efficacy endpoints will be performed on different populations (FAS set and PP set).

Unscheduled measurements and Visit 2 assessments done day(s) after the study treatment start date (therefore, not considered as baseline) will be excluded from the analysis and will be only listed.

Descriptive statistics for all efficacy analyses will be performed both on actually observed data and after the imputation of the missing data. The inferential analyses will be produced only after the imputation of the missing data.

Notably, for the purpose of the descriptive statistics (for both continuous and binary endpoints) after imputation of missing data, the average of the imputed and non-imputed data over the N samples will be calculated.

Details for the method of missing imputation are described in Appendix 12.1.

Individual subject listings for all data (including both observed and means of imputed values) will be provided.

10.1. Primary Efficacy Endpoints

The two co-primary efficacy endpoints are:

- Mean change in body weight from baseline over 12 weeks
- Mean change in 5-item Anorexia Symptom Subscale from baseline over 12 weeks

10.1.1. Primary efficacy endpoint derivation

The mean change effect for both co-primary endpoints will be computed as sum of the changes from baseline over 12 weeks by the time of the last assessment (either week 12 or before in case of death), and then divided by the number of assessments (observed or imputed) from baseline up to the time of the last assessment (either week 12 or before, in case of death).

For the derivation of the 5 item Anorexia Symptom Subscale FAACT A/CS domain at each visit see Section 4.3.1 and Appendix 12.3.

10.1.2. Primary efficacy analyses

The primary efficacy analyses will be conducted on the ITT set. Statistical hypothesis H_{0w} and H_{0a}, as defined in Section 7.3, will be tested using an Analysis of Variance (ANOVA) model including treatment group and the three stratification factors at randomization as categorical covariates, at a Type-I error threshold of 0.05 two-sided.

As anticipated in the beginning of this Section 10, inferential analysis will be done after imputation of missing data. An established number (N=100) of imputations will be

performed and for each of the imputation the ANOVA model will estimate treatment effect and the related standard error. The pooling of the estimates will be computed using the Rubin rule (1). The pooled effect and the related p-value will be interpreted for the primary efficacy analyses. In addition, the 95% confidence interval of the pooled difference and the least squares (LS)-mean with the 95% confidence interval will be presented per treatment group.

More details for the multiple imputation process are described in Appendix 12.1.

10.1.3. Sensitivity analyses

The sensitivity analyses are aimed to confirm robustness of the conclusion of the primary efficacy analyses but are not part of the multiplicity adjustment.

10.1.3.1. *Control-based imputation*

A sensitivity analysis regarding imputation of missing data will be performed using data from the placebo group only (control-based imputation) for all non-intermittent missing data. Therefore, the exceptions described in Appendix 12.1, deviating from the main control-based imputation, will not be taken into consideration.

10.1.3.2. *Fluid retention replacement analysis*

As fluid retention may possibly influence the weight increase, a sensitivity analysis will be performed for the primary endpoint “Mean change in body weight from baseline over 12 weeks”, by considering as missing the values obtained in presence of fluid retention (for the definition of fluid retention, see Section 8.8 “Definitions”).

Those missing values will be therefore imputed on the basis of the following rules for imputation:

- In case of one or more body weight assessments measured in presence of fluid retention, comprised between two assessments (pre- and post-) measured in absence of fluid retention, the lower value between the assessed weight(s) and the weight(s) estimated by the linear interpolation between pre- and post-assessments will be imputed.
- If no subsequent values observed in absence of fluid retention are available, the post-baseline weight(s) recorded in presence of fluid retention will be imputed with the baseline value if that value is lower than the recorded value. Otherwise the recorded value will be used
- If fluid retention is already present at baseline (due to protocol violation), no imputation will be performed and observed values will be analyzed.

After imputation of these values, the imputation rule for the non-intermittent missing values described in Appendix 12.1 will be applied.

Furthermore, an additional sensitivity analysis will be performed for the primary endpoint “Mean change in body weight from baseline over 12 weeks”, by excluding the patients for which fluid retention was reported at any time during the study.

10.1.3.3. *COVID-19 sensitivity analysis*

In case the subject was unwilling or unable to reach the research site or the site was not able to accommodate the visits during the COVID-19 pandemic, body weight was collected by the subject at home. Despite the recommendations to make the body weight data collection as most uniform as possible (8), high heterogeneity can be expected. This might possibly have an impact on the overall results.

A sensitivity analysis will be performed for the primary endpoint "Mean change in body weight from baseline over 12 weeks", by considering as missing the body weight values obtained when collected at home by the subject.

Those missing values will be therefore imputed on the basis of the following rules for imputation:

- In case of one or more intermittent missing value comprised between two assessments (pre- and post-) measured at site, the weight value(s) estimated by the linear interpolation between pre- and post-assessments will be imputed.
- In case of no intermittent missingness (the weight values set at missing because collected at home during COVID-19 pandemic are the last ones and no other "post" assessment is performed at site) the imputation rule for the non-intermittent missing values described in Appendix 12.1.1.2 will be applied with the following update:
    - All missing values (including also the weight values set at missing because collected at home) will be imputed assuming MNAR, specifically by means of the control-based multiple imputation method.
    - The exception 1, imputation of the missing values with the MAR approach under the circumstances described in Appendix 12.1.1.2, will be applied also on the subset of patients set at missing because collected at home

10.1.3.4. *Mis-stratification sensitivity analysis*

During the course of the study and as part of the data and medical review process, some cases of mis-stratifications (i.e.: incorrect stratification information entered into the IRT system for the purpose of the randomization of the patients) were observed and investigated with the sites involved. The final list of mis-stratifications, confirmed by the PIs, was reviewed and discussed during the BDRM.

Since the percentage of patients affected by such procedural deviations was about 7,5%, even if any imbalance due to such randomization errors is expected to be minimal, a sensitivity analysis will be performed for the co-primary endpoint on the ITT set, using the correct strata information as part of the covariates included in the statistical model. Specifically, for the 5-item Anorexia Symptom Subscale stratum (≤10 vs >10), the correct information is available in the clinical database (as part of the baseline assessments), while for the other two strata (systemic anti-cancer treatment line and cancer treatment), not being part of the clinical database variables, the correct

information to be incorporated into ADaM, displayed in the listings and used for the sensitivity analysis will be provided as external data, according to sites confirmation (see BDRM minutes and related attachment).

10.1.4. Supportive analyses

10.1.4.1. *Analyses on the FAS.*

Similarly to the primary efficacy analysis in the ITT set, the analyses will be conducted on the FAS set.

10.1.4.2. *Analysis on the PP set*

Similarly to the primary efficacy analyses, the analyses will be conducted on the PP set.

10.1.4.3. *Analysis on subgroups of the co-primary endpoint*

Similarly to the primary efficacy analyses, analyses will be conducted on the ITT Set based on the subgroups defined in Section 8.6. The statistical analyses and models will be based on the models used for the co-primary endpoints described in Section 10.1.2.

A forest plot presenting point estimates and 95% confidence intervals will be presented for weight and 5-IASS.

10.1.4.4. *Premature discontinuation and retrieved dropouts*

In order to summarize the presence of retrieved dropouts in the data of body weight and 5-IASS, a frequencies for patients who have the last efficacy assessment at the same week as the last treatment given as well as retrieved dropouts patients will be calculated. The number and percentage of patients will be presented for each category, including a sub-category of patients who have the last efficacy assessment at the same week as the last treatment given and who stayed alive during the study. Summaries will be done separately for body weight and for 5-IASS by treatment arm on the ITT set.

10.2. Secondary Efficacy Endpoints

All the analyses of the secondary endpoints will be conducted on the ITT set.

The following secondary endpoints are to be analyzed:

- Duration of treatment benefit in body weight (≥0 kg) over 12 weeks
- Duration of treatment benefit in body weight (≥ 1.5 kg) over 12 weeks
- Duration of treatment benefit in 5-item Anorexia Symptom Subscale (≥0 points) over 12 weeks

- Duration of treatment benefit in 5-item Anorexia Symptom Subscale (≥ 3 points improvement) over 12 weeks
- Mean change in FAACT 12-item A/CS domain score from baseline over 12 weeks
- Mean change in FACIT-F from baseline over 12 weeks
- Mean change in FAACT total score from baseline over 12 weeks

As described in Section 7.3 the secondary efficacy endpoints analyses will be performed by following a hierarchical order.

Subgroup analysis for the secondary endpoints, are defined in Section 8.6.

10.2.1. Secondary efficacy endpoints derivation

Secondary endpoints based on Duration of treatment benefit

The duration of treatment benefit over 12 weeks is not directly collected in the study data but should be derived from:

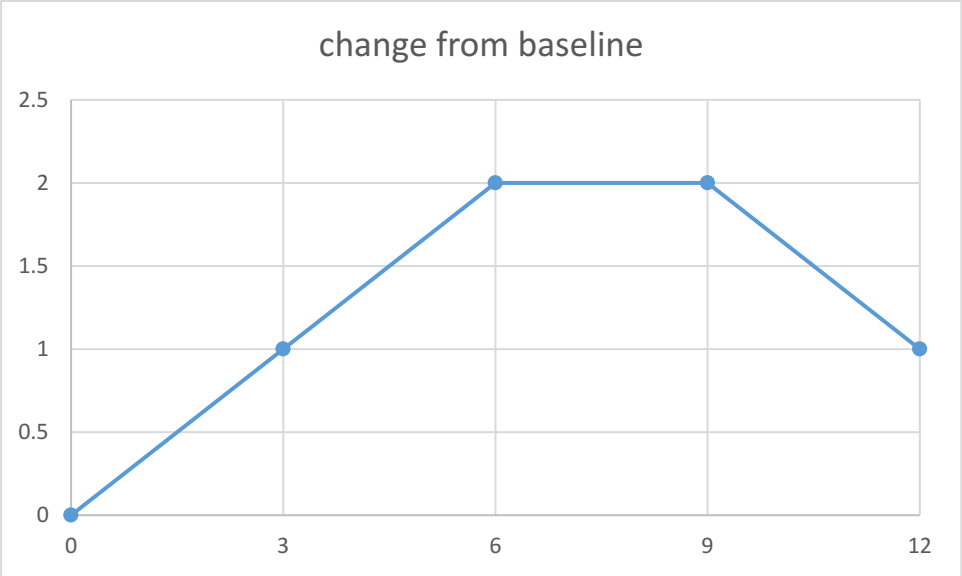
- the dates of the assessments (either observed or imputed),
- the values of the assessments (either observed or imputed)

The duration of treatment benefit over 12 weeks is measured as the period, or the sum of the periods, over 12 weeks (or less in case of death), in which the patient observed a change from baseline superior or equal to the threshold.

For the purpose of computation, it will be assumed a linear evolution between two measurements.

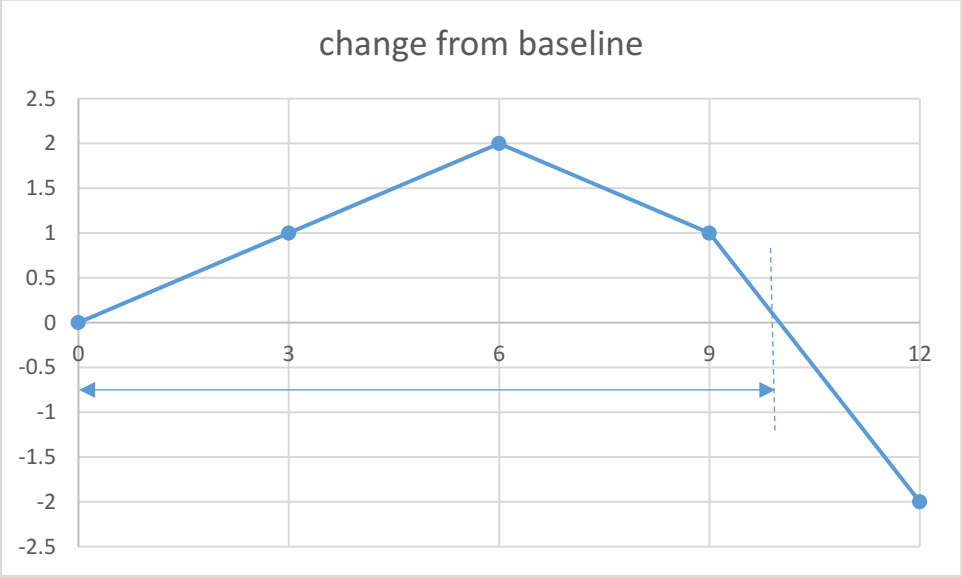
Below some examples (threshold = 0. Same approach in case of threshold=clinical meaningfulness)

Case 1: benefit all over 12 weeks. No need of linear interpolation



Duration of treatment benefit = Date of Week 12 – Date of Baseline assessment

Case 2: benefit until a given visit. Need of linear interpolation



In case 2 the algorithm should estimate the time when the line between two measurements (in the example between Week 9 and Week 12) crosses the threshold.

The equation of the line passing through two given points: (x1, y1) and (x2, y2) is the following:

(Y -Y1)/(Y2-Y1) = (X-X1) / (X2-X1)

Where Y=0 if the threshold=0 (or =clinical meaningfulness threshold, otherwise) and:

X1=date of the Visit at Week 9

Y1=value of assessment (e.g., Body Weight) at Week 9

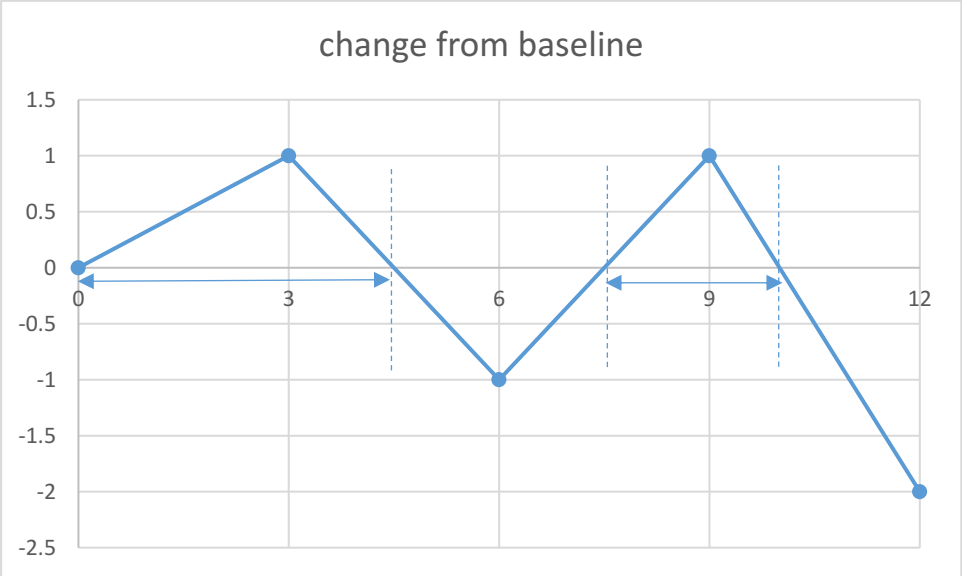
X2=date of the Visit at Week 12

Y1=value of assessment (e.g., Body Weight) at Week 12

Setting Y=0 (threshold value), the equation estimates the time X_T when the benefit ends.

The duration of benefit will be = (Date of Week 9 – Date of Baseline assessment) + (Date of X_T – Date of Week 9).

Case 3: intermittent benefit. Need of linear interpolation



For case 3 estimation of duration of benefit follows the same approach as for Case 2.

These examples describe cases with all assessments available (e.g.; after multiple imputation).

The same approach will be applied to derive the endpoints before the multiple imputation (observed values, to be presented as descriptive statistics).

If treatment benefit is not reached, the duration of treatment benefit in this patient will be set to 0; these patients will be included into the statistical analysis.

Secondary endpoints based on Mean Change from baseline over 12 weeks

The mean change effect for these secondary endpoints will be computed as sum of the changes from baseline over 12 weeks by the time of the last assessment (either week 12 or before), and then divided by the number of assessments (observed or imputed) from baseline up to the time of the last assessment (either week 12 or before, in case of death). This is aligned to the approach followed for the main analysis of the co-primary endpoints.

10.2.2. Secondary efficacy endpoints analyses

All secondary endpoints (duration of treatment benefit over 12 weeks and mean change from baseline over 12 weeks in FAACT-F 12-items, FACIT-F and FAACT total score in the secondary efficacy endpoints) will be analyzed using the same statistical model and multiple imputation approach as for the co-primary efficacy endpoints. Details are reported in Section 10.1.2 and Appendix12.1.

10.2.3. Secondary efficacy analysis on subgroups

For all secondary efficacy endpoints, analyses will be conducted on the ITT Set based on the subgroups defined in Section 8.6. The statistical analyses and models will be based on the models used for the secondary endpoints described in Section 10.2.2. For subgroups defined based on a stratification factors, the corresponding variable will be excluded from the model for the analysis.

Forest plots will be done presenting point estimates and 95% confidence intervals of each secondary efficacy endpoints.

10.3. Exploratory Endpoints

All the analyses of the exploratory endpoints will be conducted on the ITT. See Section 6.1.3 for the list of exploratory endpoints which will be analyzed.

The datasets imputed for the primary and secondary efficacy endpoints will be used also for the analysis of the exploratory endpoints.

10.3.1. Exploratory endpoints derivation.

All the rules defined for the derivation of the primary and secondary endpoints apply to the exploratory endpoints as relevant.

10.3.1.1. *Mean Change from baseline over X Weeks*

Mean Change in Body Weight, in 5-IASS and in FAACT total score from baseline over 6, 9, 15, 18, 21, and 24 Weeks are derived as the primary efficacy endpoint (see Section 10.1.1).

For instance:
Mean change in Body Weight from baseline over 24 Weeks is calculated as the sum of the changes from baseline over 24 weeks by the time of the last assessment (either week 24 or before in case of death), and then divided by the number of assessments (observed or imputed) from baseline up to the time of the last assessment (either week 24 or before, in case of death).

10.3.1.2. *Percentage of Change in Body Weight*

Percentage of change in body weight from baseline computed as
(Weight at Visit – Baseline) / Baseline * 100

10.3.1.3. *Composite Clinical Response (CCR)*

CCR is a composite measure defined as:

- 1.5 kg body weight gain from baseline and
- ≥ 3 points gain in the 5-item Anorexia Symptom Subscale score from baseline.

The two components of CCR are derived as follows:

- Achievement of 1.5 kg gain in body weight from baseline computed as
  - If change in body weight ≥ 1.5 kg then "Responder"; otherwise "Non-responder"
- Achievement of ≥ 3 points increase in 5 I-ASS from baseline computed as
  - If (5-IASS at Visit - Baseline) ≥ 3 then "Responder"; otherwise "Non-responder"

CCR is then computed as

- If "Responder" in Achievement of ≥ 1.5 kg gain in body weight, and
  If "Responder" in Achievement of ≥ 3 points increase in 5-IASS, and
  If patient is alive then "Responder";
  otherwise "Non-responder"

10.3.1.4. *Achievement of responder threshold*

At each visit:

- "Responder" if value >= clinically meaningful threshold and patient is alive
- "Non-responder" otherwise

Definition of Clinically meaningful threshold is given in Section 8.8.

10.3.1.5. *Change from baseline*

Changes from baseline to Weeks 3, 6, 9, 12, 15, 18, 21, and 24 are derived as:

("variable" at Visit – Baseline).

10.3.2. Exploratory endpoints analyses.

The exploratory efficacy analyses are aimed to explore the effect of the anamorelin on different endpoints but are not part of the Type I error multiplicity adjustment, thus have no confirmatory value.

10.3.2.1. *Mean Change from baseline over X Weeks*

The exploratory endpoints based on the "mean change from baseline over X Weeks" will be computed and analyzed like the main analysis for the primary efficacy endpoints.

10.3.2.2. *Change from baseline*

All the exploratory endpoints based on the "change from baseline" except the HAS and the PGIS/C will be analyzed using an ANCOVA model, with explanatory variables being the randomized treatment, the stratification factors and the baseline value.

The estimator of the effect will be the difference in change from baseline between the test group and the placebo group estimated as the difference in LS means of the ANCOVA model. The least squares (LS)-mean and 95% confidence interval will be presented per treatment group together with the LS-mean placebo-corrected treatment effect and its 95% confidence interval. Statistics will be evaluated in each imputed dataset and pooled using Rubin's method (1).

10.3.2.3. *Responder Analysis (CCR and Achievement of responder threshold)*

All the responder analyses (CCR and achievement of responder threshold) will be analyzed using a logistic regression, with explanatory variables being the randomized treatment, and the stratification factors. Frequency and rate of responders will be reported with Clopper-Person 95% confidence intervals (CIs). For the treatment comparison, the odds ratio will be reported with 95% CI. Statistics will be evaluated in each imputed dataset and pooled using Rubin's method (1).

10.3.2.4. *Correlations and scatter plots*

To explore the correlation between the weight gain and the improvement on the patient reported anorexia score, the Pearson's correlation coefficient between change from baseline in Body Weight and change from baseline in 5-item Anorexia Symptom Subscale respectively will be estimated overall and within each treatment group. Analysis will be performed at each Study Visit. Same analysis will be performed using the mean change over 12 weeks.

In addition, a scatter plot of the change from baseline of the two variables will be presented (change from baseline in Body Weight at Y-axis and change from baseline in 5-IASS at X-axis), with a display of the distribution of each single variable. The analysis will be performed at each Study Visit. Same analysis will be performed using the mean change over 12 weeks.

Correlation and Scatter plot will be performed both on observed data and after imputation.

10.3.2.5. *PGIS, PGIC and Hunger assessment*

The PGIS and the PGIC endpoints and the hunger assessment will not be imputed, only observed value will be summarized. PGIS and PGIC are aimed to identify the responder threshold in the 5-IASS and weight. The related statistical analyses is reported separately (6, 7).

10.3.2.6. *Change from baseline to Week 12 in body weight (categorical)*

Additionally, absolute change from baseline in body weight to Week 12 will be categorized as follows:

- <-3 kg
- < -2 kg to >= -3 kg
- < -1 kg to >= -2 kg
- <= 0 kg to >= -1 kg
- > 0 kg to <= 1 kg
- > 1 kg to <= 2 kg
- > 2 kg to <= 3 kg
- > 3 kg.

Summary will present the number and percentage of subjects for each body weight category by treatment.

11. Safety Analysis

All safety analyses will be performed on the Safety set.

11.1. Exposure

Duration of overall exposure (days) will be calculated as the difference between last tablet intake date and the first tablet intake date plus 1.

- Duration of overall exposure (days) = date of EOT - date first tablet intake +1

Duration of overall exposure will also be calculated in weeks using the following formula:

- Duration of overall exposure (weeks) = Duration of exposure (days) / 7

Duration of overall exposure, as defined above, will include any period of temporary treatment interruption.

The duration of overall exposure in days will be summarized using descriptive statistics by treatment groups.

A contingency table by treatment groups will also be presented with the following 3-weeks categories:

- [0-3)
- [3-6)
- [6-9
- [9-12)
- [12-15)
- [15-18)
- [18-21)
- [21-24)
- ≥24 weeks.

In addition, duration of actual exposure (in days and weeks respectively) will be also calculated similarly to the duration of overall exposure but removing any treatment interruption (i.e.: considering actual differences between dates of last and first tablet intake as reported in the drug accountability for each drug kit returned).

In case a patient prematurely discontinued treatment and presented an EOT date, as recorded in the related CRF form, which is considered as implausible (i.e.: EOT date greater than 25 days after the date of the first dose of the last kit), then for calculation of actual exposure the corresponding EOT date will be replaced with the date of last visit performed + 24 days.

For patients with kit(s) not returned but with information available in the drug administration CRF form about the documented tablet intake during the applicable study visit (i.e.: the study visit during which the first tablet from the kit – then never returned – was administered), this will be used to estimate the date of last intake from the kit, adding 24 days (as maximum number of tablets available in the kit to be

regularly taken) after the first intake. This approximation will be done to derive treatment interruptions for kit(s) that were not returned. This is applicable for actual exposure calculation only.

Exposure data will be listed for all subjects in the Safety Set.

11.2. Adverse Events

AEs will be coded using Medical Dictionary for Regulatory Activities (MedDRA) dictionary (version 22.0) to assign a preferred term (PT) and a system organ class (SOC) for each event.

11.2.1. Classification of Adverse Events

The reporting period for AEs is the period starting from the date/time of Informed Consent signature and lasting until Visit 11 (approximately 183 (+3) days post study drug administration on Day 1). All AEs (including SAEs) occurring within this period and not resolved by the end of the period will be documented on the eCRF as "ongoing".

Pre-Treatment AEs

An AE with onset date between date/time of Informed consent and the date of the first dose of the study drug (not included) is classified as a pre-treatment AE:

- Pre-treatment AE: AE onset date < date of the first dose of study drug;

Treatment-Emergent AEs

A treatment-emergent AE (TEAE) is an AE with onset date on or after the date of the first dose of the study drug and up to Study completion (last study visit being Visit 11: approximately Day 183 +3 days after first dose).

- TEAE: date of the first dose of study drug ≤ AE onset date ≤ date of completion

In case of treatment discontinuation an AE is classified as TEAE if:

- TEAE: date of the first dose of study drug ≤ AE onset date ≤ date of discontinuation +14 days

Notes:

- *Events where the onset date is the same as the study drug start date are assumed to be TEAEs;*
- *By protocol 14 days is the post-treatment time-window for completers between Visit 10 (last dose) and Visit 11 (follow-up by phone call).*

Any AE already present before the first dose of the study drug, that worsens in either intensity or frequency following exposure to the treatments is to be considered as new event classified as TEAE.

Post-Study AEs

All AEs fulfilling the following conditions:

- onset date > study completion or
- onset date > treatment discontinuation+14 days

are classified as post-study AE.

In case of missing or incomplete dates not directly allowing allocation to any of the three categories of AEs, see the rules defined in Section 8.3.5.

Pre-treatment AEs, TEAEs and post-study AEs will be presented separately. Summary tables will be provided on TEAE only. Pre-treatment AEs and post-study AEs will be presented in the listings only.

11.2.2. Adverse Events of Special Interest (AESI)

Adverse events of Special Interest are pre-specified as the following (PTs):

1. Aspartate aminotransferase increase (CTCAE Grade ≥ 3)
2. Alanine aminotransferase increase (CTCAE Grade ≥ 3)
3. Presyncope (any CTCAE Grade)
4. Syncope (any CTCAE Grade)
5. Ventricular arrhythmia (CTCAE Grade ≥ 3)
6. Cardiac failure (CTCAE Grade ≥ 3)
7. Sudden death
8. Seizure (any CTCAE Grade)
9. Hyperglycemia (CTCAE Grade ≥ 3)

11.2.3. Summaries of Adverse Events

A summary table reporting the total number of events, the total number of patients and the percentage of patients experiencing at least one of the following:

- Treatment-emergent adverse events (TEAEs)
- Drug-related TEAEs
- Serious TEAEs (SAEs)
- Drug-related SAEs
- TEAE CTCAE ≥ grade 3.
- TEAEs leading to treatment discontinuation
- Drug-related TEAEs leading to treatment discontinuation,
- Treatment-emergent AEs of special interest
- AEs with outcome Deaths

will be created by treatment group.

Drug-related TEAE are defined as AEs with relationship classified by the investigator as definitely, probably, possibly, unassessable or missing.

Each of the categories reported above will also be summarized in separate tables by SOC and PT:

- By decreasing frequency for SOC;

- By decreasing frequency for PT within the SOC;

Each summary table will be created by treatment group and overall.

For all adverse events, data will be prepared classifying events using MedDRA dictionary with respect to preferred term and SOC by treatment group.

Patients will be counted only once at the SOC level and will be counted once for each applicable preferred term; the total number of events will be presented for each SOC/PT.

11.2.4. Summaries of Adverse Events by Maximum CTC grade Severity

At individual patient level, the maximum CTC grade describes the worst grade experienced by a patient at any point in the treatment, irrespective to the AEs.

At a SOC/PT level, if a subject has multiple AEs with same preferred term the maximum CTC grade for that PT is used in the analysis.

If an AE has no grading, grade 3 is assigned; grade 4 or 5 will be assigned only if life-threatening consequence or death respectively will be observed.

A summary of maximum CTCAE grade at patient level will be reported by treatment group. The incidence of TEAEs in each treatment arm will also be presented by SOC and PT. For each PT the grouping by severity according to the CTCAE grade (maximum grade in case of multiple AEs with same preferred term) will be reported.

11.2.5. Summaries of Adverse Events by Maximum relationship to the study treatment

At individual patient level the highest level of relationship experienced by a patient at any point in the treatment, irrespective to the AEs, will be presented.

For summaries presenting AEs related to study drug, the number of patients with an AE, is the number of patients who have each level of relationship as the highest level of relationship for the specific PT.

11.2.6. Summaries of TEAEs of Special Interest

A table presenting details of the Treatment-emergent AEs of special interest will be provided by treatment arm. TEAESI will be also listed separately.

11.3. Laboratory Data

For continuous clinical laboratory results (hematology and biochemistry parameters), the absolute value and change from baseline will be summarized by parameter and visit using descriptive statistics for the Safety Set. For laboratory parameters where the shifts to the last assessed value are evaluated, change from baseline to the last value assessed will be summarized as well.

If the value is below the lower limit of quantification (e.g.: "< X.X"), then half the lower limit of quantification will be used in the summary tables to calculate the descriptive statistics. If the value is above the upper limit of quantification (e.g.: (e.g.: "> X.X"), then the upper limit of quantification will be used in the summary tables to calculate the descriptive statistics. In the listings, values will be listed as collected without the imputation applied.

Shift tables from baseline to the Last Value according to NCI CTCAE V5.0 grade (published on Nov 2017) will be provided for selected chemistry parameters (ALT, AST, total bilirubin, ALP, creatinine, and glucose) and hematology parameters (hemoglobin, platelet count, white blood cells, and neutrophils).

Frequency tables according to NCI CTCAE worst grade over time will be provided for ALT and AST.

In addition, % change from baseline by visit will also be measured for HbA1c and pre-albumin.

Urinalysis

Urinalysis parameters (WBC, Erythrocytes, Protein, Glucose and Bacteria) will be classified as normal or abnormal and will be summarized by parameter using the number and percentage of subjects in each category for the Safety Set.

Shift tables from baseline to the Post-baseline Values will be presented. For WBC and Erythrocytes, values classified as High or Low will be recoded to Abnormal to be presented using the same layout as for the other qualitative urinalysis parameters.

All clinical laboratory data will be listed, and values outside the normal ranges will be flagged.

All clinical laboratory values, abnormal clinical laboratory values, and clinically significant laboratory values (including data collected at any unscheduled assessments) will also be provided in data listings.

Pregnancy test

The results of pregnancy test by visit will be summarized with counts and percentages of patients by treatment in the ITT Set.

11.4. Vital Signs

Vital sign measurements for the Safety Set will be summarized using descriptive statistics by visit. The change from baseline will also be summarized in the same manner. Unscheduled vital signs assessments will be excluded from the tables.

All vital signs, weight and BMI measurements will be listed in the ITT Set including data collected at any unscheduled assessments.

In principle, any treatment emergent clinically significant change in vital signs will be reported as an AE.

11.5. Electrocardiogram (ECG)

Shift tables from baseline to the worst post-baseline results will be provided for the ECG overall interpretation. The following categories will be used: normal, abnormal and not clinically significant, abnormal and clinically significant. The frequencies of ECG overall interpretation by visit as well as the number and percentage of patients who performed the assessment will be also tabulated.

ECG measurements (HR, QT, QTcF, QTcB, PR, RR, and QRS) will be summarized using descriptive statistics by treatment group and visit. The change from baseline will also be summarized in the same manner.12-lead ECGs measurements are obtained in triplicate. Analysis will be performed using the average of the three measurements. In case of one or two measurements are missing the average will be calculated on the bases of the available measurements. Summaries showing the number and percent of patients with outlier values will be presented by treatment.

Outlier values will be identified using the following criteria:

ECG interval	Categorical outlier criteria
QTcF	Treatment-emergent value of > 450 and ≤ 480 ms when not present at baseline (new onset)
	Treatment-emergent value of > 480 and ≤ 500 ms when not present at baseline (new onset)
	Treatment-emergent value of > 500 ms when not present at baseline (new onset)
	Increase of QTcF from baseline of > 30 and ≤ 60 ms
	Increase of QTcF from baseline > 60 ms
PR	Increase of PR from baseline > 25% resulting in PR > 200 ms
QRS	Increase of QRS from baseline > 25% resulting in QRS > 120 ms
HR	Decrease of HR from baseline >25% resulting in HR < 50 bpm
	Increase of HR from baseline >25% resulting in HR > 100 bpm

Any treatment emergent clinically significant ECG abnormality will be reported as an AE.

All ECG data will be listed by patient.

11.6. Physical Examination

Any treatment emergent clinically significant physical examination abnormality will be reported as an AE. Physical examination findings will be listed.

The signs and symptoms of edema will be tabulated by visit and treatment.

11.7. ECOG Performance Status

ECOG performance status will be summarized using shift tables from baseline (Screening) to Week 9.

11.8. Overall Survival

Overall survival for the two treatment arms will be analyzed descriptively using Kaplan-Meier estimation. Efron method will be implemented for handling ties. Descriptive statistics will be provided as median time to event (and Interquartile Range). A summary of events and censored subjects will be given for each arm (n, %). K-M Plot with the survival curves for the two treatment arms will also be provided.

The difference in survival between anamorelin and placebo, descriptively assessed using a stratified Cox regression analysis (stratified by stratification factors listed in Section 8.4), will be provided as estimate of hazard ratio (HR) and 95%CIs.

Event or censor time in weeks will be calculated as follows:

- Event time (i.e. death) = ([date of death] – [Study Treatment Start Date] +1 day)/7
- Event time (i.e. no death) = ([last follow-up contact date – [Study Treatment Start Date] +1 day)/7

In this derivation, “Study Treatment Start date” instead of “Randomization date* is used because Overall Survival is considered as Safety variable.

11.9. Tumor Assessments

Change in tumor assessment at the latest CT scan performed during the treatment period compared to CT scan at baseline (overall response categorized into complete response, partial response, stable disease and progressive disease) will be summarized by treatment group with frequencies and percentages.

Baseline CT scan is performed at screening visit or within 28 days of the first dose of study drug.

All patients at all sites will perform CT scans as per standard local practice and responses evaluated according to the RECIST 1.1. Description of target and non-target lesions as well as overall response before randomization (if any) will be provide as listing.

12. APPENDICES

12.1. Missing value imputation process

Imputation of missing values will be performed following:

- the multiple imputation (MI) method for missing data, based on the approach described in literature by Ratitch (1, 2).
- Single imputation method, which is described, case by case, in the process reported below.

Conventions:

- All observed data will be included in any analysis;
- The missing data of the variables being part of the primary efficacy endpoints (body weight, 5-item Anorexia Symptom Subscale score) and the secondary/exploratory efficacy endpoints will be imputed;
- A specific sensitivity analysis for the primary efficacy endpoint “Mean change in body weight from baseline over 12 weeks" will be performed to replace the body weight value of the patients with fluid retention at the date of the assessment. These values will be firstly set to missing and, then, imputed with the imputation approach reported in Section 10.1.3.2.
- No imputation will be performed for survival status and no data will be imputed after patient death.

When a measurement will be missing for patients who are alive, the imputation will depend on the pattern of missingness**: non-monotone** vs. **monotone**:

- In case of non-monotone pattern, missing values may occur in the middle of a subject record (e.g., missed intermediate visit). These are also called “Intermittent missing values”.
- In case of a monotone pattern, it is possible to arrange variables so that, for all subjects, missing values are always located in a block at the end of data records (e.g., due to study discontinuation). These are also called “non-Intermittent missing values”.

For the definition see also Section 8.8.

12.1.1. Imputation rules

12.1.1.1. *Imputation for Intermittent missing values (non-monotonic missingness)*

These values will be imputed with the value obtained by linear interpolation between prior and post values.

12.1.1.2. *Imputation rules for Non-intermittent missing values (monotonic missingness)*

Imputation rules for non-intermittent missing values are based on a main rule and some exceptions.

The rule and the exceptions will be applied separately for each efficacy parameter.

Main rule

All missing values will be imputed assuming MNAR, specifically by means of the control-based multiple imputation method.

Reason: Most of the patients are expected to have missing values from time of treatment stop. In addition, the patients under study are characterized by poor prognosis and the reasons for treatment discontinuation are expected to be more related to the background disease than to treatment failure. For this reason, it cannot be assumed that the values after the treatment discontinuation would follow the same trend as before discontinuation because the general conditions, with specific exceptions described below, after discontinuation get to worsen. The implementation of the control-based imputation is conservative because it assumes that the deterioration of the patient condition affects mainly the patients in the active treatment group.

Exceptions

1. Independent from reason for discontinuation of treatment, if:
   - at least two non-missing post-baseline assessments are available (first assessment = X_1 and last assessment = X_{last}) and,
   - $X_{last} - X_1 \geq 0$ and $X_{last} \geq$baseline,

   then impute all the missing values with the MAR multiple imputation approach including randomised treatment as the covariate.
2. If reason for discontinuation of treatment = “Physician decision” or = “Withdrawal by subjects” and:
   - at least two non-missing post-baseline assessments are available (first assessment = X1 and last assessment = X_{last}) and,
   - $X_{last} - X_1 < 0$ and $X_{last} <$baseline,

   then impute all missing values to the worst value observed for the subject, considering only the on-treatment (post-baseline) timepoints.

Reasons:

Imputation for Exception 1

This exception is applicable for any reason for discontinuation, when the trend reveals an improvement of the efficacy assessments over time, with the last post-baseline assessment (X_{last}) better than both the baseline and the first post-baseline assessment (X_1). Based on these conditions, it is assumed that the decision of stopping the treatment is not related to the treatment itself (since the efficacy parameter was improving over time). Due to the improvement over time, it can also be assumed that

the treatment effect would have been maintained as it was before interruption for the rest of the observation period. For this reason, the missing values after the treatment discontinuation will be imputed with the Missing at Random (MAR) multiple imputation approach including randomized treatment as the covariate. This approach might be slightly optimistic for values missing at visits far from the visit of the last assessment, however this will not impact the primary endpoint that is limited to week 12, considering that at least assessments at week 3 and at week 6 are requested by the rule.

Imputation for Exception 2

This exception is limited to the reasons of treatment discontinuation that are most likely related to lack of efficacy of the treatment; the criterion for considering the unknown values as related to lack of efficacy is the negative trend of the observed efficacy assessments over time, as documented by the last post-baseline assessment (X_{last}) worse than the baseline and the first post-baseline assessment (X_1). Based on these conditions, it is assumed that the decision of stopping the treatment can be related to the treatment itself (lack of efficacy). The missing values will be imputed with the worst value observed for the patient considering only the on-treatment (post-baseline) timepoints.

Final revision of the above rules was done at the BDRM and they have been confirmed.

The details of the process are showed in the following steps.

STEP 1: Single/Multiple imputations

- Impute all intermittent missing data
- Create 100 records for each value
- Application of the Main Rule:
  Apply MNAR Control-Based imputation on all subjects (placebo and active treatment) with at least one missing assessment.
  - MI on all non-intermittent missing
  - only placebo patients contribute to the imputation.

  Expectation after imputation: only 1 dataset including:
  - ANAM and placebo patients without missing data (all assessments observed)
  - ANAM and placebo patients with at least 1 missing assessment imputed by the imputation process.

  Note: it is theoretically possible that aberrant values are derived because of the imputation. Aberrant values are defined as follows:
  - For each item of the questionnaires:
    a) Imputed value < the lower limit (0)
    b) Imputed value > the upper limit (4)
  - For the body weight:
    a) Imputed value < the lower limit (30 Kg)
    b) No upper limit for the Body Weight is fixed.

In the cases (a) the imputed value will be set equal to the lower value.

In the cases (b) the imputed value will be set equal to the upper value.

Unscheduled assessments of body weight and questionnaires are not considered for imputations. These values will be listed only (and have all ANL0XFLs set to blank).

- Application of Exception 1:
  Apply MAR (as correction to the previous control-based imputation on patients fulfilling Exception 1)
  Note: Values to be imputed with MAR need first to be re-set at missing.
  A flag before starting with the imputation process should identify the cases where the exception 1 should be applied.
  Imputed values after MAR imputation should be within the aberrant values described above for the MNAR imputation
- According to Exception 2:
  Impute worst value at subject level.
  A flag before starting with the imputation process should identify the cases where the exception 2 should be applied. Worst value should be calculated before any imputation.
- The imputation process does not distinguish between missing due to treatment discontinuation and death. All imputed assessments after death must be set to missing.

No covariates, apart from treatment arm (as by variable or covariate, as relevant) and baseline value will be included in the MI models.

For each variable, 100 imputed values will replace each missing value.

The following variables will be separately imputed according to the above mentioned single or multiple imputation methods:

- Weight
- 5-IASS
- 4-item Anorexia Concerns Subscale (4-IACS)
- 12-item A/CS domain
- FAACT total score
- FACT-G total score
- FAACT TOI
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)
- FACIT-F TOI score

The PGIS and the PGIC endpoints and the hunger assessment will not be imputed.

Once the imputation process is complete, the dataset will contain the observed values and the imputed values.

All the values imputed after the intercurrent event death and set to missing, in respect to the “while-alive” principle will be considered as non-responders in the “responder analyses”.

The value of the derived endpoints (e.g.: mean change, duration of benefit, change from baseline) will be computed once the imputation datasets are created.

STEP 2: Perform the relevant analyses.

Each of the imputed datasets for primary/secondary/exploratory efficacy endpoints will be analyzed separately using the method declared in the proper section (for instance, the primary efficacy analysis, see Section 10.1.2, will be based on an ANOVA model to compare the two treatment groups, at a Type-I error threshold of 0.05 two-sided).

Each model will provide an estimate of the treatment effect and of the related standard error.

STEP 3: Combine Parameter Estimates.

The pooling of the estimates will be computed using the Rubin rule (1). The pooled effect and the related p-value will be interpreted for the primary (secondary/exploratory) efficacy analyses.

12.2. Statistical Software

Statistical analyses will be performed using SAS Version 9.4 (or higher) for Windows.

The SAS procedures and data steps presented below are only examples of the procedures to be implemented.

The SEEDS are the ones to be used in the procedures.

12.2.1. Single/Multiple imputation

Notes:

- Intermittent missing data are imputed by linear interpolation, single imputation
- N=100 imputation samples (variable “_IMPUTATION_”) are created prior to next step (PROC MI)
- TRTPN=’0’: refers to placebo treatment arm

12.2.1.1. *Placebo (control-based) imputation*

```
PROC MI DATA=DATA1 SEED=191124 OUT=DATA2A NIMPUTE=1;
      BY _IMPUTATION_ ;
      CLASS TRTPN;
      VAR W_00 &VARS;
      MONOTONE REGRESSION;
      MNAR MODEL(&VARS
                  / modelobs= (TRTPN='0'));
RUN;
```

Macrovariable &VARS includes W_03 for imputing data at visit 3 weeks, includes W_03 and W_06 for imputing data at visit 6 weeks, and so on.

Note: No "Where" clause is used. The model processes ALL records. For the records without any missing data the model does not apply any action.

12.2.1.2. *MAR imputation*

```
PROC MI DATA=DATA1 SEED=458122 OUT=DATA2B  NIMPUTE=1;
      BY _IMPUTATION;
      CLASS TRTPN;
      VAR TRTPN W_00 W_03 W_06 W_09 W_12 W_15 W_18 W_21 W_24;
      MONOTONE REGRESSION;
RUN;
```

Note: DATA2B dataset derives from DATA2A (see output dataset from MNAR placebo control-based imputation) after re-setting at missing the assessments as per Exception 1 (see above)

12.2.1.3. *Worst value imputation*

```
DATA DATAx ; SET DATAx;
Where &sel;
<<identification of the worst value (WORST) for the subject where:
At least 2 non-missing:
Xlast<X1
Xlast<baseline >>
Run;

Data Datax2;
Set datax;
      if W_09=. then W_09=WORST;
      if W_12=. then W_12=WORST;
      if W_15=. then W_15=WORST;
      if W_18=. then W_18=WORST;
      if W_21=. then W_21=WORST;
      if W_24=. then W_24=WORST;
RUN;
```

&sel = selection of reason for discontinuation in ("withdrawal by subject", "physician decision")

12.2.1.4. *Combining the imputations*

Imputations will follow the sequence as specified in Section 12.1.1.2.

Values imputed after a patient died will be re-set to missing (which guarantees to have no response after death, as a missing value in SAS is below any value, including the threshold).

Imputed values that are outside the range of the parameter will be set to the relevant limit of the range.

12.2.2. Data Analysis

Each of the imputations will be analyzed by usual SAS procedure.

For the primary efficacy analyses and for the other analyses based on ANOVA, the pooling will be done as follows:

```
/** Model by imputation **/
proc mixed data=[imputed dataset] ;
      by  _IMPUTATION_ ;
      class TRTPN(ref="0") [classCovariates];
      model meanchange_w = [COVARIATES] TRTPN ;
      lsmeans TRTPN / cl diffs ;
      estimate 'Anamorelin - Placebo' trtpn 1-1 /cl;
      ods output LSMEANS=MIlsmeans;
      ods output estimates=MIancova;
run ;
```

12.2.3. Pooling of the analysis

The results of the analysis will be summarized with SAS ® PROC MIANALYZE.

```
/** perform the pooling **/
proc mianalyze data= MIancova all ;
      modeleffects estimate ;
      stderr stderr ;
      ods output parameterestimates= ANCOVADIFFS_OUT;
run ;

The output dataset MIlsmeans can be used to obtain the LSmeans by treatment,
by using the relevant proc myanalyze, as follows:

proc sort data=MIlsmeans; by trtpn;      where effect='trtpn'; run;
proc mianalyze data=MIlsmeans_&var.;
    modeleffects Estimate ;
    stderr stderr ;
    ods output parameterestimates=ANCOVAls;
        by trtpn;
run;
```

For the Responder analyses the following approach will be implemented.

```
      *From Ratitch's paper 'Combining Analysis Results from Multiply
Imputed Categorical Data' PharmaSUG 2013 - Paper SP03;
      ***Log-transform odds ratio estimates;
       data lgodds_t;
          set lgodds (where=(index(effect,"TREATMENT")));
            or_lg=log(oddsratioest);
            or_lg_se   =(log(uppercl)-log(lowercl))/(2*1.96);
      run;

      ***Combine transformed estimates;
      PROC MIANALYZE DATA=lgodds_t;
        ods output parameterestimates=or_lg;
```

```
  modeleffects or_lg;
  stderr       or_lg_se;
run;

***Back-transform combined values;
data or;
  set or_lg;
  OR_pool  =EXP(ESTIMATE);             *Pooled Odds Ratio;
  OR_pool_L=OR_pool*EXP(-1.96*STDERR); *Pooled lower limit;
  OR_pool_U=OR_pool*EXP( 1.96*STDERR); *Pooled lower limit;
run;
```

12.3. Questionnaires scoring guidelines

FAACT Scoring Guidelines (Version 4) – Page 1

Instructions:* 1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.
4. Add subscale scores to derive total scores (TOI, FACT-G & FAACT).
5. **The higher the score, the better the QOL.**

Subscale	Item Code	Reverse item?		Item response	Item Score
PHYSICAL WELL-BEING (PWB)	GP1	4	-	_____	=_____
	GP2	4	-	_____	=_____
	GP3	4	-	_____	=_____
	GP4	4	-	_____	=_____
Score range: 0-28	GP5	4	-	_____	=_____
	GP6	4	-	_____	=_____
	GP7	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=**PWB subscale score**

SOCIAL/FAMILY WELL-BEING (SWB)	GS1	0	+	_____	=_____
	GS2	0	+	_____	=_____
	GS3	0	+	_____	=_____
	GS4	0	+	_____	=_____
Score range: 0-28	GS5	0	+	_____	=_____
	GS6	0	+	_____	=_____
	GS7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____=**SWB subscale score**

EMOTIONAL WELL-BEING	GE1	4	-	_____	=_____
	GE2	0	+	_____	=_____

(EWB)	GE3	4	-	_____	=_____
	GE4	4	-	_____	=_____
Score range: 0-24	GE5	4	-	_____	=_____
	GE6	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 6: _____
Divide by number of items answered: _____**=EWB subscale score**

FUNCTIONAL WELL-BEING (FWB)	GF1	0	+	_____	=_____
	GF2	0	+	_____	=_____
	GF3	0	+	_____	=_____
	GF4	0	+	_____	=_____
Score range: 0-28	GF5	0	+	_____	=_____
	GF6	0	+	_____	=_____
	GF7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____**=FWB subscale score**

FAACT Scoring Guidelines (Version 4) – Page 2

Subscale	Item Code	Reverse item?		Item response	Item Score
ANOREXIA CACHEXIA SUBSCALE (ACS)	C6	0	+	_____	=_____
	ACT1	0	+	_____	=_____
	ACT2	4	-	_____	=_____
	ACT3	4	-	_____	=_____
	ACT4	4	-	_____	=_____
Score range: 0-48	ACT6	4	-	_____	=_____
	ACT7	4	-	_____	=_____
	ACT9	4	-	_____	=_____
	O2	4	-	_____	=_____
	ACT10	4	-	_____	=_____
	ACT11	4	-	_____	=_____
	ACT13	0	+	_____	=_____

Sum individual item scores:_____
Multiply by 12: _____
Divide by number of items answered: _____**=AC subscale score**

To derive a FAACT Trial Outcome Index (TOI):
Score range: 0-104

TOI _____ + _____ + _____ =_____**=FAACT**
(**PWB** score) (**FWB** score) (**ACS** score)

To Derive a FACT-G total score:
Score range: 0-108

_____ + _____ + _____ + _____=_____=**FACT-G Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score)

To Derive a FAACT total score:
Score range: 0-156

_____ + _____ + _____ + _____ + _____
=_____=**FAACT Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score) (**ACS** score)

*For guidelines on handling missing data and scoring options, please refer to the Administration and Scoring Guidelines in the manual or on-line at www.facit.org.

The following scores are also derived from the FAACT measure:

- 5-item Anorexia Symptom Subscale (5-IASS): obtained by summing the 5 anorexia symptom reversed items scores within the ACS subscale (i.e., "good appetite", "interest in food drops", "food tastes unpleasant", "get full quickly", and "difficulty eating rich/heavy foods"); the range of possible scores is 0-20.
- 4-item Anorexia Concerns Scale: obtained by summing the 4 anorexia concern reversed items scores within the ACS subscale (i.e., "amount I eat sufficient", "worried about weight", "concerned about thinness", and "pressured to eat"); the range of possible scores is 0-16.

The following scores will be computed:

- 5-IASS = C6+(4-ACT3)+(4-ACT6)+(4-ACT7)+(4-ACT10)
- 4-item Anorexia Concerns Subscale = ACT1+(4-ACT2)+(4-ACT4)+(4-ACT9)

FACIT-Fatigue Subscale Scoring Guidelines (Version 4) – Page 1

Instructions: * 1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.
4. **The higher the score, the better the QOL.**

Subscale	Item Code	Reverse item?		Item response	Item Score
FATIGUE SUBSCALE	HI7	4	-	_____	=_____
	HI12	4	-	_____	=_____
	An1	4	-	_____	=_____
	An2	4	-	_____	=_____
	An3	4	-	_____	=_____

Score range: 0-52

An4	4	-	_____	=_____
An5	0	+	_____	=_____
An7	0	+	_____	=_____
An8	4	-	_____	=_____
An12	4	-	_____	=_____
An14	4	-	_____	=_____
An15	4	-	_____	=_____
An16	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 13: _____
Divide by number of items answered: _____**=Fatigue Subscale score**

Additionally, based on the here-above Fatigue Subscale (FACIT-F), the Physical wellbeing and the Functional well-being, the FACIT-F TOI will be computed as: FACIT-F TOI = FACIT-F + PWB + FWB; the range of possible scores is 0-108.

12.4. Validation and QC Plan

The program reviewer is responsible for reviewing each project program and output associated with deliverable product. Program logs will be reviewed for logical, syntax and fatal errors. The review in SAS will include, but is not limited to, all ERRORS, WARNINGS, BY-VALUE merge messages, NOTES, and UNINITIALIZED variables. Program logs will also be reviewed for accurate and consistent variable and observation counts following each procedure and data step.

A quality control (QC) statistician will be responsible for checking and reviewing the work produced using whatever method he/she feels is appropriate (e.g., SAS code review, hand calculation, etc.) to confirm the quality of the output.

Outputs will be reviewed for typographical errors, misspellings and spurious values or results and to check the consistency with the SAP. Outputs will be cross-checked against each other for accuracy and consistency. This procedure will include comparisons of subject group numbers, counts of subjects at each observation point, and consistency of results for variables between outputs.

Findings of the QC reviews will be communicated to the party responsible for making necessary changes. The programs will be QC reviewed again after modifications.

After final review, when no further changes are required to produce the deliverable, the program reviewer and QC statistician will produce documentation to indicate that they have successfully performed all of their responsibilities.

12.5. Mock Tables

The mock TLFs are included into a separate document and have been reviewed and agreed along with the SAP text.

13. CHANGES IN THE PLANNED ANALYSIS FROM STUDY PROTOCOL

The following changes from study protocol have been considered:

Efficacy Analysis

- Sensitivity Analysis:
  - a subset of patients performed visits remotely due to COVID-19 restrictions and some important assessment (Body Weight) was collected and reported by the patients during the phone call. A sensitivity analysis not foreseen in the study protocol has been added in the SAP, by setting at missing the impacted values and managing these missing values using the Multiple Imputation process.
  - as the presence of fluid retention may affect the assessment of weight, an additional sensitivity analyses (other than the one already anticipated in the protocol) will be conducted for the primary endpoint "Mean change in body weight from baseline over 12 weeks, excluding entirely patients with fluid retention observed at any point during the study.
  - some cases of mis-stratifications were confirmed for a subset of patients. Thus, a sensitivity analysis has been added in the SAP to repeat the primary efficacy analysis using the correct strata information as part of the covariates included in the statistical model.
- Exploratory Analysis:
  - with the aim to better describe the clinical benefit and the responsiveness to anamorelin, a separate analysis (Changes from baseline to Week 3, 6, 9, 12, 15, 18, 21, and 24) on the most important patient-reported symptoms and functional items extracted from FAACT total score have been added:
    - Items GP1, GP7, GF5, GF6 GF7 related to physical and functional well-being
    - Items ACT1, ACT2, ACT4, ACT13 related to A/CS part
    - Items HI7, HI12, AN2, AN5, AN12 related to FACIT-F (Fatigue) questionnaire
  - the evaluation of the effect of anamorelin on Overall Survival has been incorporated into the Study Design as a safety objective to verify that no detrimental effect of anamorelin were observed. Anamorelin is not an anticancer medication, and as such, it is not expected to have a measurable positive impact on PFS or OS. The ANAM studies were not designed to detect a potential survival benefit of anamorelin mediated by body weight gain and improvement of anorexia symptoms. The relationship between changes in body weight and the 5-IASS will therefore not be analyzed, neither in the overall study population nor at

the level of treatment arms. Composite Clinical Response (CCR) definition was updated using final clinically meaningful thresholds for body weight gain (1.5 kg) and 5-IASS increase (3 points) resulted from the quantitative analysis of a separate study (7), combined with evidence from the embedded qualitative interview sub-study (6)

- With the aim to present a complete picture of the primary endpoint, the absolute change from baseline in body weight to Week 12 will also be presented for some pre-defined categories agreed during BDRM.

- Subgroup Analysis (Section 8.6): due to the inadequate number of patients within each subgroup, as described during the BDRM, the subgroups analysis (analysis of the coprimary and secondary efficacy endpoints in each stratum defined by the variable) will not be performed for the following variables:
  - Race,
  - Ethnicity
  - Country (Geographic region)
  - Patients with at least one treatment emergent adverse event with preferred term related to fluid retention and/or patients with signs of edema.

Safety Analysis

- Laboratory Data: for Hematology and Blood Chemistry descriptive statistics based on absolute value and change from baseline by parameter and visits have been added. This analysis has been added because by protocol shift tables for only a subset of Hematology and Blood Chemistry parameters were planned.
- Overall Survival: an analysis, based on the stratified Cox regression model has been planned, in addition to the pre-planned analysis based on Kaplan-Meier estimates.
- Tumor assessments: The best overall tumor response summary as planned in the protocol has been replaced by the last overall response. This is because anamorelin is not expected to have an impact on tumor and the best response was finally not considered of safety interest.

14. REFERENCES

1. Ratitch B. Multiple Imputations. In: O'Kelly M and Ratitch B, editors. Clinical Trial with Missing Data: A Guide for Practitioners. West Sussex: John Wiley & Sons Ltd, 2014. p 185-256.

2. Ratitch B, Lipkovic I, and O'Kelly M. Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG 2013 - Paper SP03. 2013.

3. Yang Yuan, SAS Institute Inc. Sensitivity Analysis in Multiple Imputation for Missing Data. SAS Proceedings, SAS Global Forum 2014: Paper SAS270-2014

4. Clinical Study Report HT-ANAM-301 - ROMANA 1 - Anamorelin HLC in the treatment of Non-Small Cell Lung Cancer-Cachexia (NSCLC-C): a randomized, double-blind, placebo-controlled, multicenter, Phase 3 study to evaluate the safety and efficacy of anamorelin HCL in patients with NSCLC-C.

5. Clinical Study Report HT-ANAM-302 - ROMANA 2 – Anamorelin HLC in the treatment of Non-Small Cell Lung Cancer-Cachexia (NSCLC-C): a randomized, double-blind, placebo-controlled, multicenter, Phase 3 study to evaluate the safety and efficacy of anamorelin HCL in patients with NSCLC-C.

6. Qualitative Embedded Interviews among Phase 3 Clinical Trial Patients with Advanced Non-small Cell Lung Cancer (NSCLC) and Malignancy-associated Weight Loss and Anorexia. EVA-31327-01 | April 28, 2023| Version 3.0.

7. Responder Definitions for 5-Item Anorexia Symptoms Subscale and Body Weight in Patients with Advanced Non-small Cell Lung Cancer and Malignancy-associated Weight Loss and Anorexia. Results Memo. EVA-31327-00-001 | 8 May 2023 | Version 5.0.

8. COVID-19 Subject Remote Visit Instruction_version 1.0_FINAL_06 April 2020