I8F-MC-GPGU Statistical Analysis Plan Version 1

A Randomized, Placebo-Controlled, Parallel-Arm Study to Investigate the Effect of Once-Weekly Tirzepatide on Energy Expenditure and Food Intake in Obese Subjects

NCT04081337

Approval Date: 14-Jan-2022

1. Statistical Analysis Plan
I8F-MC-GPGU: A Randomized, Placebo-Controlled, Parallel-Arm Study to Investigate the Effect of Once-Weekly Tirzepatide on Energy Expenditure and Food Intake in Obese Subjects

LY3298176 for Obesity

Phase 1, single-center, randomized, sponsor-, investigator- and subject-blind, placebo-controlled, parallel-arm study, in obese subjects to compare tirzepatide 15 mg to placebo.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I8F-MC-GPGU
Phase 1

Statistical Analysis Plan electronically signed and approved by Lilly on date provided below.

Approval Date: 14-Jan-2022 GMT

2. Table of Contents

Section **Page**

Table of Contents

Tables **Page**

Table of Contents

Figure **Page**

3. Revision History

This is statistical analysis plan (SAP) Version 1 and approved prior to the database lock.

4. Study Objectives

Table GPGU.4.1 shows the objectives and endpoints of the study.

Table GPGU.4.1. Objectives and Endpoints

Objectives	Endpoints
Primary To assess the effect of tirzepatide versus placebo on SMR after 18 weeks (±1 week) of treatment	• The change from baseline to Week 18 in SMR, as measured in the respiratory chamber (whole-room calorimetry)
Secondary To assess the effect of tirzepatide versus placebo after 18 weeks (±1 week) of treatment: • Appetite and food intake	• The change from baseline to Week 18 in appetite VAS • The change from baseline to Week 18 in food intake as assessed by ad libitum meal test
• Energy expenditure	• The change from baseline to Week 18 in 24-hour EE
• Substrate oxidation rates	• The change from baseline to Week 18 in 24-hour RQ • The change from baseline to Week 18 in sleep RQ • The change from baseline to Week 18 in duration of periods with RQ<0.80 • The change from baseline to Week 18 in fat, protein, and carbohydrate oxidation
• Body weight and body composition	• The change from baseline to Week 18 in BW • The change from baseline to Week 18 in body fat-free mass • The change from baseline to Week 18 in body fat mass • The change from baseline to Week 18 in the percentage of body fat mass
• Lipid metabolism	• The change from baseline to Week 18 in triglycerides, cholesterol, LDL, VLDL, and HDL cholesterol, FFA, glycerol, 3-hydroxybutyrate, and acylcarnitines assessed during sMMTT • The change from baseline to Week 18 in ApoB-48, ApoB-100, ApoC-III, and LPL during sMMTT • The change from baseline to Week 18 in fasting concentration of leptin, adiponectin, IGFBP 1 and 2
• Insulin sensitivity	• The change from baseline to Week 18 in fasting insulin resistance (as measured by the HOMA2 method [HOMA2-IR]) and postprandial insulin sensitivity indices (Matsuda, OGIS, and Stumvoll) assessed during sMMTT

Objectives	Endpoints
• Glucose control	• The change from baseline to Week 18 in fasting and postmeal glucose during sMMTT • The change from baseline to Week 18 in hemoglobin A1c
Exploratory To assess the effect of tirzepatide versus placebo after 18 weeks (±1 week) of treatment: • GIPR signaling, lipid metabolism, carbohydrate metabolism, and insulin signaling pathways in subcutaneous adipose tissue	• The change from baseline to Week 18 in markers of GIPR signaling, lipid metabolism, carbohydrate metabolism, and insulin signaling pathways as measured in exploratory mRNA expression, lipidomics, metabolomics, or targeted protein assays of subcutaneous adipose tissue biopsy samples, receptor expression, signal transduction, and metabolic changes in adipose tissue
• Safety and tolerability	• Adverse events • Safety laboratory parameters • Frequency of treatment-emergent anti-tirzepatide antibodies • Treatment-emergent depression and suicidal ideation as assessed using C-SSRS and PHQ-9

Abbreviations: ApoB-48 = apolipoprotein B-48; ApoB-100 = apolipoprotein B-100; ApoC-III = apolipoprotein C-III; BW = body weight; C-SSRS = Columbia-Suicide Severity Rating Scale; EE = energy expenditure; FCI = Food Craving Inventory; FPQ = Food Preference Questionnaire; FFA = free fatty acid; GIPR = glucose-dependent insulinotropic polypeptide receptor; HDL = high-density lipoprotein; HOMA2 = Homeostatic Model Assessment of Insulin Resistance; HOMA2-IR = insulin resistance as measured by the HOMA2 method; IGFBP = insulin-like growth factor binding protein; LDL = low-density lipoprotein; LPL = lipoprotein lipase; mRNA = messenger ribonucleic acid; OGIS = Oral Glucose Insulin Sensitivity; PHQ-9 = Patient Health Questionnaire-9; RQ = respiratory quotient; sMMTT = standardized mixed-meal tolerance test; SMR = sleep metabolic rate; VAS = visual analog scale; VLDL = very low-density lipoprotein.

5. Study Design

5.1. Study Design and Treatment

The study GPGU is a Phase 1, single-center, randomized, sponsor (including study team members directly involved in study management)-, investigator- and subject-blind, placebo-controlled, parallel-arm study in obese subjects. This study is designed to assess differences in the mechanisms of action with respect to energy balance, lipid and carbohydrate metabolism, and insulin sensitivity between tirzepatide and placebo, both in combination with a low-calorie diet in non-diabetic, obese subjects.

The primary objective of this study is to compare the 2 treatment groups (tirzepatide and placebo) for change from baseline in sleep metabolic rate (SMR) assessed in a respiratory chamber (whole-room indirect calorimetry) after 18 weeks (±1 week) of treatment. Secondary objectives will assess the effect of tirzepatide versus placebo after 18 weeks (±1 week) of treatment on appetite and food intake, energy expenditure, substrate oxidation rates, body weight and body composition, lipid metabolism, insulin sensitivity and glucose control.

The study will consist of the following periods: 4-week screening period, 2-week lead-in period, 18-week (±1 week) treatment period, and a 4-week safety follow-up period. Subjects will be screened within 4 weeks prior to lead-in. Eligible subjects will enter lead-in (Visit 2) within 2 weeks prior to randomization. Subjects will be randomized in a 1:1 ratio to tirzepatide or placebo on Day -1. Sponsor (study team members directly involved in study management), investigators and subjects will be blinded to tirzepatide and placebo treatment.

Tirzepatide dosing will start at a dose of 2.5 mg QW for 2 weeks, followed by a step-wise dose escalation to 5 mg QW for 2 weeks, and 10 mg QW for 4 weeks, followed by 15-mg QW dose which is then maintained for the remainder of the treatment period (10 weeks ±1 week). Subjects will return to the clinical research unit (CRU) every week for QW dose administration for 18 weeks (±1 week).

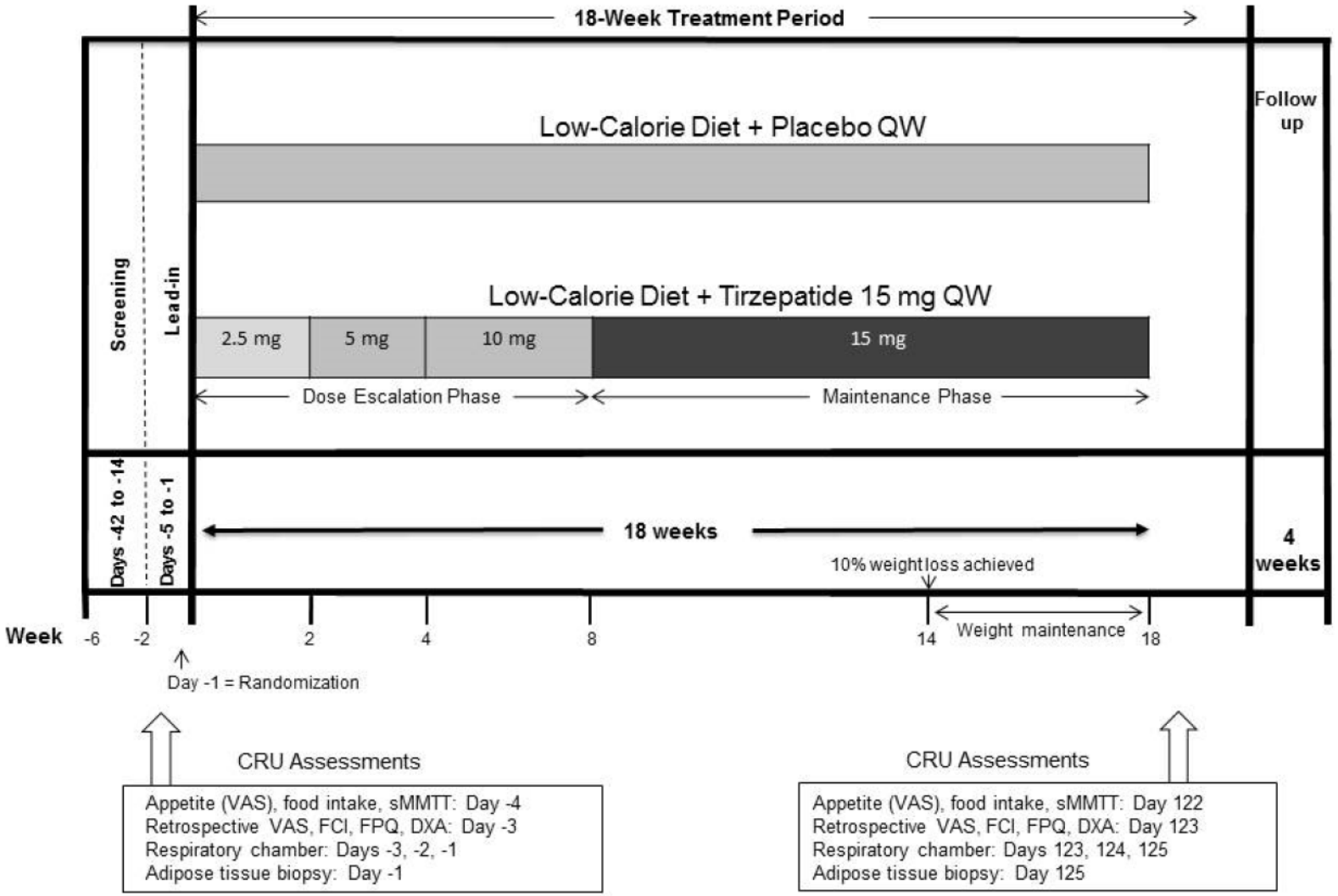
Subjects in both groups will receive a standardized dietary intervention to target a body weight (BW) reduction of 10% (±2%) after approximately 14 weeks of treatment. From Week 14 to the end of treatment period at Week 18 (±1 week), dietary intervention will shift to promoting maintenance of 10% weight loss within target range. Subjects will be readmitted to the clinical research unit (CRU) at Week 18 on Day 121 to perform end-of-treatment study procedures. These CRU study procedures will be performed in all subjects, irrespective of whether they reached their BW target or not, and according to the same schedule as during the lead-in period.

Figure GPGU.5.1 illustrates the study design.

Figure GPGU.5.2 illustrates the clinical research unit procedures.

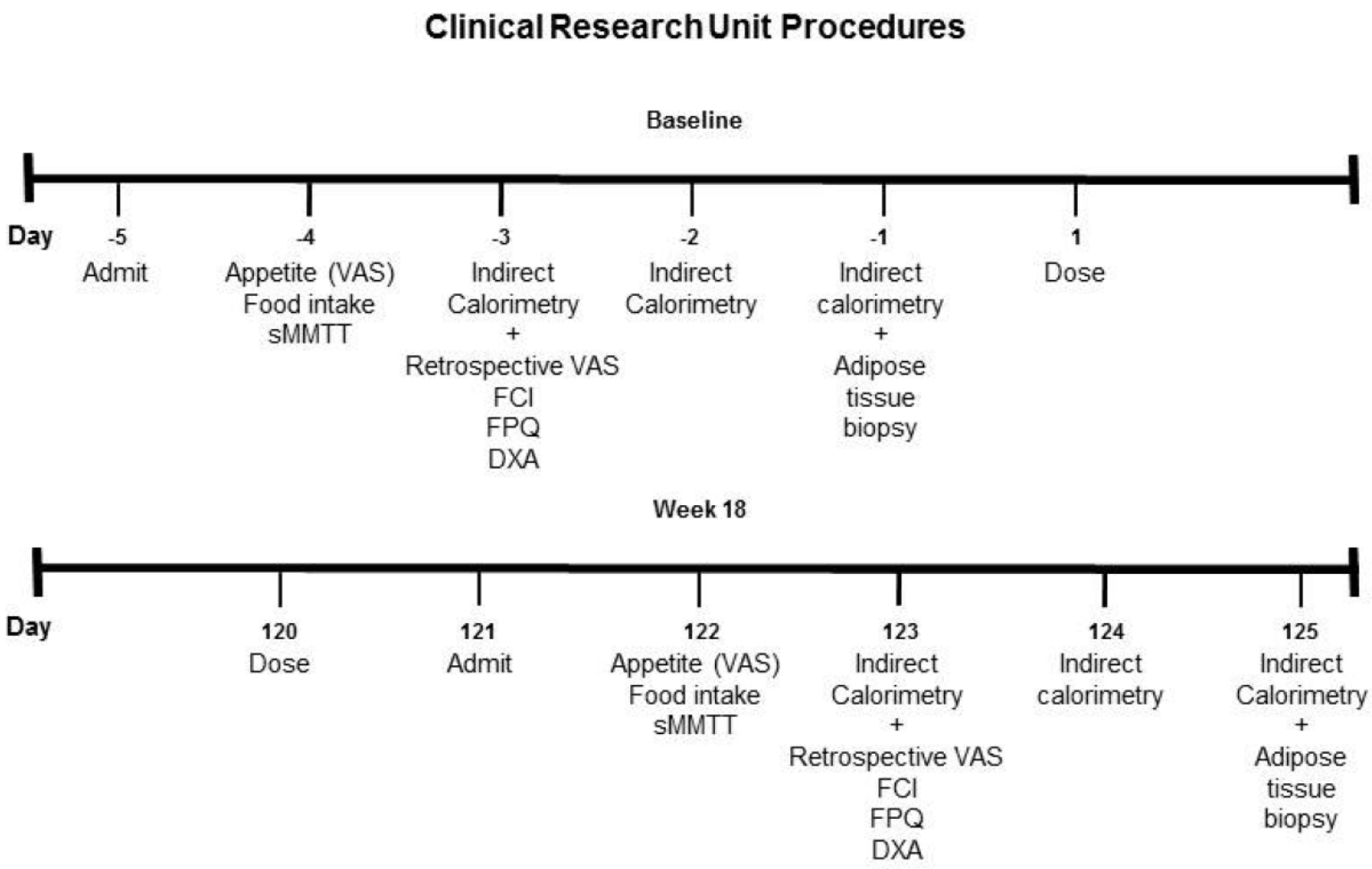
5.2. Determination of Sample Size

Up to 56 subjects are planned to be randomized so that approximately 46 subjects complete the study assuming 15% discontinuation rate. These 56 subjects will be randomized to QW tirzepatide or placebo in a 1:1 ratio. The estimated standard deviation of the change in SMR from baseline is 117 kcal/day (Heilbronn et al. 2006). With an expected treatment difference of 100 kcal/day, this provides at least 80% power for the comparison of tirzepatide versus placebo based on 2 sample t-test using 2-sided test at alpha level of 0.05 for the assessment of the primary endpoint.



Abbreviations: DXA = dual energy X-ray absorptiometry; FCI = Food Craving Inventory; FPQ = Food Preference Questionnaire; QW = once weekly; sMMTT = standardized mixed-meal tolerance test; VAS = visual analog scale.

Figure GPGU.5.1. Illustration of study design for Protocol I8F-MC-GPGU.



Abbreviations: DXA = dual energy X-ray absorptiometry; FCI = Food Craving Inventory; FPQ = Food Preference Questionnaire; sMMTT = standardized mixed-meal tolerance test; VAS = visual analog scale.

Figure GPGU.5.2. Illustration of clinical research unit procedures.

6. A Priori Statistical Methods

6.1. Population for Analyses

For the purpose of analysis, Table GPGU.6.1 defines 3 analysis sets.

Table GPGU.6.1. Analysis Populations/Data Sets

Population/Data Set	Description
All randomized population	All subjects who are randomly assigned a treatment arm.
Safety population	All randomized subjects who are exposed to at least 1 dose of study drug (tirzepatide or placebo), regardless of whether they completed all protocol requirements.
PD analysis set *	Evaluable PD data from all randomized subjects who are exposed to at least 1 dose of study drug.

Abbreviations: PD = pharmacodynamic.

* Protocol deviations will be considered for their severity/impact and will be taken into consideration whether subjects should be excluded from PD analysis set.

6.2. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly) or its designee. Any change to the statistical methods described in the protocol will require a protocol amendment only if it changes a principal feature of the protocol. Any other change to the statistical analyses and the justification for the change will be documented in SAP.

Unless otherwise specified, safety analyses will be conducted on the safety population (Table GPGU.6.1), and pharmacodynamic (PD) analyses will be conducted on the PD analysis set (Table GPGU.6.1).

Unless otherwise specified, baseline is defined as the last scheduled non-missing measurement collected during Visit 1 (screening) and Visit 2 (lead-in) before first dose.

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and the confidence interval (CI) will be calculated at 95%, 2-sided.

Unless otherwise specified, Week 18 represents Week 18-visit (Visit 20).

6.3. Study Participant Characteristics

The following subject baseline demographic and clinical characteristics will be summarized for the safety population by study treatment (including overall):

- age (years)
- sex (male, female)
- race
- ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- body weight (kg) (Day -1, lead in)
- height (cm)

- body mass index (BMI) (kg/m^2)
- BMI group (<30*, ≥30 to <35, ≥35 to <40, and ≥40 kg/m^2)
  * If no inadvertent enrollment in BMI <30 kg/m^2 group, this category can be removed.
- waist circumference (cm)

6.4. Study Participant Disposition

A listing of subject disposition for the randomized population will be provided. Subject disposition will be summarized by treatment.

A listing of subjects who discontinue from the study for any reason for the randomized population will be provided, and the extent of their participation in the study will be reported. The reason for their discontinuation from study will be reported. The listing will also include age, sex, and race.

6.5. Concomitant Therapy

Concomitant medication will be listed and summarized by treatment and World Health Organization Drug Dictionary Medication Class and Preferred Term for the safety population.

6.6. Treatment Exposure and Compliance

6.6.1. Treatment Exposure

The duration of exposure to study medication (tirzepatide or placebo) is defined as:

Date of last dose of study treatment - date of first dose of study treatment +7 days

The duration of exposure to study medication will be listed and summarized by treatment using the safety population.

6.6.2. Treatment Compliance

Treatment compliance will be assessed using the safety population. Treatment compliance for each patient will be listed, including the percent compliance for 18-week treatment period taking into account early termination. Percent compliance and treatment compliance will be summarized by treatment.

The percent compliance will be calculated as:

$$(Number\ of\ injections\ administered\ [regardless\ of\ actual\ dose\ in\ mg\ administered] \div total\ number\ of\ injections\ expected\ to\ be\ administered) \times 100$$

Patients will be considered treatment compliant if taking ≥75% of their scheduled doses for 18-week treatment period taking into account early termination (specifically excluding the period after early termination).

When assessing treatment compliance, the missed doses and interrupted doses will be taken into consideration as described in protocol (Section 7.2.1 Selection and Timing of Doses and Section 8.1.2 Temporary Interruption of Investigational Product).

6.7. Important Protocol Deviations

Important protocol deviations (IPDs) are defined as deviations from the study protocol that may significantly compromise the data integrity and/or patients' safety. The details of identification of IPDs are provided in a separate document (ie, the trial issue management plan). A listing/table of IPDs will be provided by the study manager after database lock.

6.8. Pharmacokinetic Analyses

6.8.1. Pharmacokinetic Parameter Estimation

Sparse pharmacokinetic (PK) samples will be collected across the 18-week (±1 week) treatment duration. Tirzepatide concentrations will be determined to support an understanding of tirzepatide exposure over the treatment duration to compare with expected tirzepatide PK based on historical understanding.

6.8.2. Pharmacokinetic Statistical Inference

The PK samples will be collected predose on baseline (Day 1), Day 29, Day 57, Day 120, and at follow-up visit. Summary table of PK concentration over time will be provided. No PK parameters will be derived; thus no statistical analyses of PK parameters are planned.

6.9. Pharmacodynamic Analyses

6.9.1. Pharmacodynamic Parameter Estimation

In this study, pharmacodynamic parameters will be used to assess the mechanisms of action of study treatments on sleep metabolic rate, energy expenditure, substrate oxidation rates, appetite visual analog scale (VAS) scores, food intake, lipid metabolism, insulin sensitivity, glucose control, food-intake related questionnaires FCI and FPQ, and body composition. The planned assessments will be performed at baseline (lead in) and postbaseline visits (including at the end of the treatment period at Week 18) (see Table GPGU.6.2 for details).

Table GPGU.6.2. A Summary of Pharmacodynamic Assessment-Related Study Procedures

Day -4 (Lead in) Day 122 (Week 18)	Day -3 (Lead in) Day 123 (Week 18)	Day -3 to -1 (Lead in) Day 123 to 125 (Week 18)

• During sMMTT (early morning)
  ○ Appetite VAS
  ○ Blood samples
• Ad libitum food intake test (lunch and dinner)
	• FCI and FPQ questionnaires
	• Retrospective VAS
	• DXA scan (body composition)
		• Respiratory chamber (approximately 2×23 hours, 8:00am – 7:00am)

Abbreviations: DXA = dual energy X-ray absorptiometry; FCI = food craving inventory; FCQ = food preference questionnaire; sMMTT = standardized mixed-meal tolerance test; VAS = visual analog scale.

* The DXA scans may be scheduled at baseline (lead-in) any time of Day -5 to Day -1 and at Week 18 any time of Day 123 to Day 125. Retrospective VAS is additionally administered weekly from Week 1 through Week 15. FCI and FPQ questionnaires are additionally administered on Day 50 (Week 8). Adipose tissue biopsy (abdominal) samples are collected on Day -1 (lead in) and Day 125 (Week 18). HbA1c is collected on screening, Day -4 (lead in), Day 57 (Week 9), Day 122 (Week 18), and safety follow-up visit. Body weight is measured on screening, Day -1 (lead in), and weekly from Week 2 to Week 17, and Day 120, Day 123, Day 124, Day 125.

Pharmacodynamic measurements collection

Respiratory Chamber (Day -3 to -1 [lead in]; Day 123 to 125 [Week 18])

Subjects will enter the respiratory chamber (whole room calorimetry) at approximately 8:00 am on Day -3, -2 and Day 123, 124, and be removed at the end of each 23-hour measurement period. Total carbon dioxide production (VCO_2) and total oxygen consumption (VO_2) will be sampled and measured at 1-minute intervals over 23 hours in a respiratory chamber (whole-room calorimeter). This will be used to calculate energy expenditure (EE), respiratory quotient (RQ), and substrate oxidation rates. Activity counts will be recorded using a motion detector for the entire 23-hour measurement period and will be used to determine activity driven effects on EE and RQ.

Note: 24-hr energy expenditure (kcal/day), protein oxidation (g/day), fat oxidation (g/day), and carbohydrate oxidation (g/day) are calculated based on 23-hour measurement period and extrapolated to 24 hours. Energy expenditure and RQ will also be calculated for event (eg sleep, rest) period, and energy expenditure will be extrapolated to 24 hours. The chamber measurements of Day -3, Day -2, Day 123, Day 124 will be provided in datasets to statisticians/programmers.

The study site uses the formulas below during the calculation:

- Energy expenditure
  - EE [kcal/min] = 3.941× VO_2 [L/min] + 1.106 × VCO_2 [L/min] – 2.17 * urinary nitrogen [g/min] (Weir 1949)
  - Sleep metabolic rate [kcal/24hr] = EE during sleep extrapolated to 24 hours
- Respiratory quotient
  - 24-hour RQ = total VCO_2 [L/24hr] / total VO_2 [L/24hr]
  - Sleep RQ = VCO_2 during sleep / VO_2 during sleep

- Substrate Oxidation (Frayn 1983; Jéquier et al. 1987)
  - 24-hr protein oxidation [g/24hr] = 6.25 × urinary nitrogen [g/24hr]
  - 24-hr fat oxidation [g/24hr] = 1.689 × total VO_2 [L/24hr] – 1.689 × total VCO_2 [L/24hr] – 0.324 × protein oxidation [g/24hr]
  - 24-hr carbohydrate Oxidation [g/24hr] = 4.113 × total VCO_2 [L/24hr] – 2.907 × total VO_2 [L/24hr] – 0.375 × protein oxidation [g/24hr]

Body Composition (DXA Scan; lead in [any time of Day -5 to Day -1] and Week 18 [any time of Day 123 to Day 125])

The dual energy X-ray absorptiometry (DXA) scans (GE iDXA™ whole-body scanner) will be conducted at baseline (lead-in, any time of Day -5 to Day -1) and Week 18 (any time of Day 123 to Day 125) to determine body composition. The DXA scan will be performed in duplicate at each scheduled time point. The DXA measurements from the 2 tests on the same day will be collected by the study site and provided in datasets to statisticians/programmers.

During sMMTT (Day -4 [lead in], Day 122 [Week 18]; early morning)

The key objectives of the sMMTT are to assess α and β cell function and indirectly assess insulin sensitivity under physiological conditions. The sMMTT will be initiated between 7:00 to 7:30 am with the subject in a seated position. The start of the meal will be defined as time point 0 (zero). Subjects will consume a High Protein Boost® over a 5-minute period maximum. Measurements are collected over the next 4 hours.

Appetite VAS during sMMTT

- Appetite VAS (at 5 time points): -10, 60, 120, 180, 240 minutes

Blood sampling during sMMTT

- Glucose (serum):
  - at 11 time points: -10, -1, 15, 30, 60, 90, 120, 150, 180, 210, 240 minutes
- Insulin (serum):
  - at 9 time points: -10, -1, 15, 30, 60, 90, 120, 180, 240 minutes
- 3-hydroxybutyrate (ie beta-hydroxybutyrate), acylcarnitines, free fatty acid (FFA), and glycerol:
  - at 6 time points: -10, 60, 90, 120, 180, 240 minutes
- Lipid panel:
  - including triglycerides, total cholesterol, low-density lipoprotein cholesterol (LDL-C), very low-density lipoprotein cholesterol (VLDL-C), high-density lipoprotein cholesterol (HDL-C), apolipoprotein B-48 (ApoB-48), apolipoprotein B-100 (ApoB-100), apolipoprotein C-III (ApoC-III), and lipoprotein lipase (LPL):
    - at 6 time points: -10, 60, 90, 120, 180, 240 minutes
- Leptin, adiponectin, insulin-like growth factor binding protein (IGFBP) 1 and 2 and IGFBP 2 sample:
  - at -10 minute

Ad Libitum Food Intake Test (Day -4 [lead in], Day 122 [Week 18]; lunch and dinner)

Food intake will be quantified through ad libitum food-intake tests at lunch and dinner at:

- Day -4 (lead in), Day 122 (Week 18)

Other PD Measurements

Food Craving Inventory (FCI)

- Day -3 (lead in), Day 50 (Week 8), and Day 123 (Week 18)

Food Preference Questionnaire (FPQ)

- Day -3 (lead in), Day 50 (Week 8), and Day 123 (Week 18)

Retrospective VAS

- Administered 16 times: Day -3 (lead in), weekly from Week 1 through Week 15 (Day 1, Day 8, Day 15, Day 22, Day 29, Day 36, Day 43, Day 50, Day 57, Day 64, Day 71, Day 78, Day 85, Day 92, Day 99), and Week 18 (Day 123).

Lipid panel (Including triglycerides, total cholesterol, LDL-C, VLDL-C, HDL-C, ApoB-48, ApoB-100, ApoC-III, and LPL)

3-hydroxybutyrate (ie beta-hydroxybutyrate), acylcarnitines, FFA, and glycerol

Leptin, adiponectin, IGFBP 1 and IGFBP 2

- Day -4 (lead in, 10 minutes prior to sMMTT [fasting]), Day 29 (Week 5; fasting), Day 57 (Week 9; fasting), Day 122 (Week 18; 10 minutes prior to sMMTT [fasting])

HbA1c

- Screening, Day -4 (lead in), Day 57 (Week 9), and Day 122 (Week 18), safety follow-up visit

AUC Calculation

All area under the concentration versus time curve (AUC) measures will be calculated using the trapezoidal rule. For example, if $\{x_k\}$ is a partition of the desired time interval [a,b], where measurements at each x_k is denoted as y_k and $a = x_0 < x_1 < \cdots < x_{N-1} < x_N = b$.

- Total AUC (trapezoidal rule) is defined as

$$AUC_{[a,b]} = \sum_{k=1}^{N} \frac{(y_k + y_{k-1})(x_k - x_{k-1})}{2}$$

- Incremental AUC measures use the time zero value relative to the AUC measured unless otherwise specified, ie

$$incremental\ AUC_{[a,b]} = AUC_{[a,b]} - y_0 \times (b - a)$$

Note:

1. If any lab measurements above are
    a. below the quantification limit (eg <QL), ½ × QL will be used instead for calculation;
    b. above the quantification limit (eg >QL), 1.1 × QL will be used instead for calculation,

    if deemed appropriate after review of data.
2. Unless otherwise specified, in case of no planned collection at 0 minute (fasting) for blood samples:
    - For pre-sMMTT serum glucose and serum insulin, the average of non-missing measurements at -1, -10 minutes will be used as the value for 0 minute to calculate AUC.
    - For pre-sMMTT 3-hydroxybutyrate, acylcarnitines, FFA, and glycerol, lipid panel, the measurement at -10 minute will be used as the value for 0 minute to calculate AUC.
3. For $AUC_{[a,b]}$, if measurement of any timepoints between time a and b (ie $a < t < b$) is missing, then this timepoint will not be used in the calculation. AUC will be calculated based on non-missing measurement and timepoints. If measurements at time a and b are missing, then AUC will be missing.

6.9.2. Pharmacodynamic Statistical Inference

Unless specified otherwise, PD parameters at baseline and post baselines, as well as change from baseline values, will be summarized by treatment, and the change from baseline to Week 18 will be analyzed.

6.9.2.1. Primary Pharmacodynamic Analyses

Sleep Metabolic Rate

The primary endpoint in the study is the change from baseline (Day -3 to -1) to Week 18 (Day 123 to 125) in sleep metabolic rate (SMR; energy expenditure during sleep period) for comparison of tirzepatide 15 mg versus placebo, which is measured in the respiratory chamber (whole-room calorimetry).

The primary endpoint will be analyzed using analysis of covariance (ANCOVA) to compare the effect of tirzepatide versus placebo at Week 18 on SMR (kcal/day):

- Independent variables: treatment as a factor, baseline SMR (kcal/day), change from baseline to Week 18 in fat free mass, and change from baseline in fat mass as covariates.
- Response variable: change from baseline to Week 18 in SMR.
- Where Fat Mass = DXA Region Fat (%)/100 * DXA Weight
- Where Fat Free Mass = DXA Weight – Fat Mass.

Baseline SMR will be analyzed with analysis of variance (ANOVA) with independent variable, treatment as a factor.

Note: Energy expenditure is calculated for sleep period and extrapolated to 24 hours. The calculated sleep metabolic rate (sleep EE in Event Data) will be provided in datasets to statisticians/programmers. The average value of nonmissing measurements at Days -3, Day -2 (two 23-hr sessions) will be used as value for baseline, and the average value of nonmissing measurements at Days 123, Day 124 will be used as value for Week 18.

Inferential statistics include least squares (LS) means and standard error of SMR for tirzepatide versus placebo, the estimated treatment difference (tirzepatide - placebo), and corresponding 2-sided 95% CI, and associated p-value.

Sensitivity analysis on primary endpoint will be conducted in a similar manner on the body weight target population (ie subjects who reach the body weight target range of -10% ±2% [-8% to -12%, inclusive] at Week 18 from baseline and have evaluable PD data).

Additional analyses will be conducted with ANCOVA with

- Independent variables: treatment as a factor, baseline SMR (kcal/day), change from baseline to Week 18 in DXA weight as covariates
- Response variable: change from baseline to Week 18 in SMR

Additional analyses may be conducted as follows:

- Step 1: conduct ANCOVA model with
  - Independent variables: baseline fat free mass, baseline fat mass, age, sex
  - Response variable: baseline SMR
  - Where Fat Mass = DXA Region Fat (%)/100 * DXA Weight
  - Where Fat Free Mass = DXA Weight – Fat Mass.
- Step 2: with estimated model from Step 1 to predict SMR at Week 18 using
  - Week 18 fat free mass, Week 18 fat mass, age, sex
- Step 3: calculate:
  $$magnitude\ of\ residual\ SMR\ at\ Week\ 18\ (metabolic\ adaptation\ ) = SMR\ measured\ at\ Week\ 18 - SMR\ predicted\ at\ Week\ 18\ from\ Step\ 2$$
- Step 4: conduct ANOVA on metabolic adaptation at Week 18 from Step 3 to compare tirzepatide to placebo

6.9.2.2. Secondary Pharmacodynamic Analyses

The following measurements are secondary endpoints in the study to compare the effect of tirzepatide 15 mg versus placebo after 18 weeks (±1 week) of treatment.

Appetite and Food Intake

Secondary PD measures to assess the effect of tirzepatide 15 mg versus placebo on appetite and food intake:

- change from baseline to Week 18 in appetite VAS assessed during sMMTT for
  - 4 individual ratings: hunger, fullness, satiety, prospective food consumption
  - 4 individual ratings: desire for 4 specific foods (sweet, salty, savory, and fatty)
  - overall appetite score

- all at 5 time points during sMMTT respectively: -10 (fasting), 60, 120, 180, 240 minutes

Note: The aim of the appetite VAS is to determine the effects of study treatments on appetite sensations and desire for specific foods.

The VAS scales will be analyzed as continuous variables on the 0 to 100 scale for individual components:

- For the above first 4 individual ratings, 0 = Not at all, 100 = Extremely
- For the last 4 individual ratings on desire for specific foods, 0 = Yes, very much, 100 = No, not at all

All 8 individual VAS scores (for each component) will be documented in electronic Case Report Form (eCRF). Overall appetite score is calculated as the average of the 4 individual scores (Flint et al. 2000; Flint et al. 2013; van Can et al. 2014) as follows:

overall appetite score = (satiety + fullness + [100 - prospective food consumption] + [100 - hunger]) / 4

The higher overall appetite score indicates less appetite, and the lower score indicates more appetite.

The above appetite VAS scores parameters (overall and 8 individual scores) will be analyzed using ANCOVA to compare the effect of tirzepatide versus placebo at each of 5 timepoints during sMMTT, respectively.

- Independent variables: treatment as a factor, baseline as a covariate
- Response variable: change from baseline to Week 18.

Time profile plots of appetite VAS parameters (mean ± standard deviation) during sMMTT (over 5 time points) will be provided by treatment for baseline and Week 18.

- change from baseline to Week 18 in food intake assessed by ad libitum meal test at
    - lunch
    - dinner
    - lunch and dinner combined

Each for the following 11 food intake variables:

- total food intake
    - g (the amount of grams of food consumed)
    - kcal (energy intake)
- carbohydrate
    - g
    - kcal
    - percent of energy (carbohydrate [kcal]/energy intake [kcal])
- protein
    - g

  - kcal
  - percent of energy (protein [kcal]/energy intake [kcal])
- fat (saturated and unsaturated)
  - g
  - kcal
  - percent of energy (fat [kcal]/energy intake [kcal])

Note: Food intake will be quantified separately for lunch, dinner, and combined for lunch and dinner. The amount of grams of fat, carbohydrates and protein consumed during lunch and during dinner will be documented in eCRF. There are 4 kcal in a gram of carbohydrate or protein, and 9 kcal in a gram of fat will be used to calculate energy intake related variables.

The above food intake parameters (at lunch, at dinner, and at lunch and dinner combined) will be analyzed using ANCOVA to compare the effect of tirzepatide versus placebo:

- Independent variables: treatment as a factor, baseline as a covariate
- Response variable: change from baseline to Week 18

Energy Expenditure (EE)

Secondary PD measures to assess the effect of tirzepatide 15 mg versus placebo on parameters of EE:

- change from baseline to Week 18 in 24-hour EE

The baseline 24-hr EE will be analyzed with ANOVA with independent variable, treatment as a factor.

At Week 18, the 24-hour EE will be analyzed with 2 ANCOVA models (similar to primary endpoint SMR)

- Model 1:
  - Independent variables: treatment as a factor, baseline 24-hr EE (kcal/day), change from baseline to Week 18 in fat free mass, and change from baseline in fat mass as covariates.
  - Response variable: change from baseline to Week 18 in 24-hr EE.
  - Where Fat Mass = DXA Region Fat (%)/100 * DXA Weight
  - Where Fat Free Mass = DXA Weight – Fat Mass.
- Model 2 (additional analyses):
  - Independent variables: treatment as a factor, baseline 24-hr EE (kcal/day), change from baseline to Week 18 in DXA weight as covariates
  - Response variable: change from baseline to Week in 24-hr EE

Additional analyses may be conducted as follow:

- Step 1: conduct ANCOVA model with
  - Independent variables: baseline fat free mass, baseline fat mass, age, sex
  - Response variable: baseline 24-hr EE
  - Where Fat Mass = DXA Region Fat (%)/100 * DXA Weight

  - Where Fat Free Mass = DXA Weight – Fat Mass.
- Step 2: with estimated model from Step 1 to predict 24-hr EE at Week 18 using
  - Week 18 fat free mass, Week 18 fat mass, age, sex
- Step 3: calculate:
  $magnitude\ of\ residual\ 24hr\ EE\ at\ Week\ 18$
  $= 24hr\ EE\ measured\ at\ Week\ 18 - 24hr\ EE\ predicted\ at\ Week\ 18\ from\ Step\ 2$
- Step 4: conduct ANOVA on residual 24-hr EE at Week 18 from Step 3 to compare tirzepatide to placebo

Substrate Oxidation Rate

Secondary PD measures to assess the effect of tirzepatide 15 mg versus placebo on parameters of substrate oxidation rates:

Respiratory quotient

- change from baseline to Week 18 in 24-hour respiratory quotient (RQ; ratio)
- change from baseline to Week 18 in sleep RQ (ratio)
- change from baseline to Week 18 in duration of periods with RQ<0.80 (minute)

The above RQ parameters will be analyzed with ANCOVA model

- Independent variables: treatment as a factor, baseline RQ parameter value as a covariate
- Response variable: change from baseline to Week 18 in RQ parameter

Protein, fat, carbohydrate oxidation

- change from baseline to Week 18 in protein oxidation (g/day)
- change from baseline to Week 18 in fat oxidation (g/day)
- change from baseline to Week 18 in carbohydrate oxidation (g/day)

The above oxidation rates parameters will be analyzed with

- Step 1: calculate adjusted oxidation parameters
  Adjusted oxidation rate (g/day) = oxidation rate (g/day) / 24-hr EE (kcal/day) * 1000 (kcal/day)
- Step 2: ANCOVA model
  - Independent variables: treatment as a factor, baseline adjusted oxidation rate as a covariate
  - Response variable: change from baseline to Week 18 in adjusted oxidation rate parameters

Note: 24-hr energy expenditure, RQ, protein oxidation, fat oxidation, and carbohydrate oxidation are calculated based on 23-hour measurement period and extrapolated to 24 hours. RQ will also be calculated for sleep period and extrapolated to 24 hours. The calculated 24-hr EE, 24 hr RQ, sleep RQ (in Event Data), duration of period with RQ<0.8, and 24-hr substrate oxidation (protein, fat, carbohydrate oxidation) will be provided in datasets to statisticians/programmers.

The average value of nonmissing measurements at Days -3, Day -2 (two 23-hr sessions) will be used as value for baseline, and the average value of nonmissing measurements during Days 123, Day 124 will be used as value for Week 18.

Body Weight and Body Composition

Secondary PD measures to assess the effect of tirzepatide 15 mg versus placebo on parameters of body weight, and body composition using DXA scan:

- change from baseline to Week 18 in
  - body weight (scale weight; kg)

Note: Body weight is measured twice at each visit, and the average of non-missing measurements will be used for analyses.

Body weight was measured at baseline Day -1 (lead in) and multiple times post baseline. Plots of mean body weight and mean change from baseline values over time will be provided by treatment. The change from baseline to Day 8, Day 15, Day 22, Day 29, Day 36, Day 43, Day 50, Day 57, Day 64, Day 71, Day 78, Day 85, Day 92, Day 99, Day 106, Day 113, and Day 120-Day 125 (average of nonmissing measurements at Days 120, 123, 124, 125 used) will be analyzed using a mixed-model repeated measures (MMRM) method with restricted maximum likelihood (REML) estimation.

- Independent variables: treatment, time, and treatment-by-time interaction as fixed effects, baseline body weight as a covariate, and patient as a random effect
- Response variable: change from baseline in body weight

An unstructured covariance structure will be used to model the relationship of within-subject errors. If this model fails to converge, the following variance covariance structures will be tested in order until convergence is achieved:

- heterogeneous Toeplitz
- heterogeneous first order autoregressive
- heterogeneous compound symmetry
- Toeplitz
- first order autoregressive, and
- compound symmetry.

The first covariance structure that converges will be used.

- change from baseline to Week 18 in
  - body fat-free mass (kg)
  - body fat mass (kg)
  - percentage of body fat mass (total region % of fat)

Note: DXA scan parameters: DXA Weight and DXA Region Fat (%) will be provided in datasets to statisticians. Fat mass and fat free mass will be calculated for 2 DXA scans on the same day:

- Fat Mass = DXA Region Fat (%)/100 * DXA Weight

- Fat Free Mass = DXA Weight – Fat Mass

The mean value of nonmissing measurements from the 2 DXA scans on the same day will be calculated for analyses.

The above body composition parameters will be analyzed using ANCOVA to compare the effect of tirzepatide versus placebo:

- Independent variables: treatment as a factor, baseline body composition parameter as a covariate
- Response variable: change from baseline to Week 18 in body composition parameter.

Lipid Metabolism

Secondary PD measures to assess the effect of tirzepatide 15 mg versus placebo on parameters of lipid metabolism:

- change from baseline to Week 18 in lipid metabolism measurements during sMMTT in:
  - triglycerides
  - cholesterol
  - LDL cholesterol
  - VLDL cholesterol
  - HDL cholesterol
  - FFA
  - glycerol
  - 3-hydroxybutyrate (ie beta-hydroxybutyrate)
  - acylcarnitines

All the above lipid metabolism lab measurements will be calculated for (1) fasting; (2) total $AUC_{0\text{-}4hr}$; (3) incremental $AUC_{0\text{-}4hr}$ for analyses.

Note: For fasting measurements, it is collected on Day -4 (lead in, 10 minutes prior to sMMTT), Day 29 (Week 5), Day 57 (Week 9), Day 122 (Week 18; 10 minutes prior to sMMTT). For AUC parameters, it is derived for baseline and Week 18 during sMMTT.

Note: See Section 6.9.1 for AUC calculation details. All these fasting lab measurements and AUCs will be reported in both International System of Units (SI Units) and conventional units.

- change from baseline to Week 18 in lab measurements during sMMTT of:
  - ApoB-48
  - ApoB-100
  - ApoC-III
  - LPL

All the above lipid metabolism lab measurements will be calculated for (1) fasting; (2) total $AUC_{0\text{-}4hr}$; (3) incremental $AUC_{0\text{-}4hr}$ for analyses.

Note: For fasting measurements, it is collected Day -4 (lead in, 10 minutes prior to sMMTT), Day 29 (Week 5), Day 57 (Week 9), Day 122 (Week 18; 10 minutes prior to sMMTT). For AUC parameters, it is derived for baseline and Week 18 during sMMTT.

Note: See Section 6.9.1 for AUC calculation details. All these fasting lab measurements and AUCs will be reported in both SI Units and conventional units.

The above available laboratory fasting parameters will be analyzed in a manner similar to the secondary PD analysis on body weight using MMRM to compare the effect of tirzepatide versus placebo.

- Independent variables: treatment, time, and treatment-by-time interaction as fixed effects, baseline as a covariate, and patient as a random effect
- Response variable: change from baseline

The above available laboratory AUC parameters during sMMTT will be analyzed using ANCOVA to compare the effect of tirzepatide versus placebo:

- Independent variables: treatment as a factor, baseline as a covariate
- Response variable: change from baseline to Week 18.

Time profile plots of the lab measurements (mean ± standard deviation) during sMMTT (over 6 time points) will be provided by treatment for baseline and Week 18.

- change from baseline through Week 18 in fasting concentration of:
  - leptin
  - adiponectin
  - IGFBP 1
  - IGFBP 2

The above fasting leptin, adiponectin, IGFBP 1 and 2 will be analyzed in a manner similar to the secondary PD analysis on body weight using MMRM since the lab tests are scheduled to be measured at baseline (Day -4) and 3 times postbaseline (Day 29 [Week 5], Day 57 [Week 9] and Day 122 [Week18]).

- Independent variables: treatment, time, and treatment-by-time interaction as fixed effects, baseline as a covariate, and patient as a random effect
- Response variable: change from baseline

Note: If any lab measurements above are (1) below the quantification limit (eg <QL), ½ × QL will be used instead for calculation; (2) above the quantification limit (eg >QL), 1.1 × QL will be used instead for calculation, if deemed appropriate after review of data. All these lab fasting measurements will be reported in both SI Units and conventional units.

Insulin Sensitivity

Secondary PD measures to assess the effect of tirzepatide 15 mg versus placebo on insulin sensitivity:

- change from baseline to Week 18 in insulin sensitivity indices assessed during sMMTT:
  - **HOMA2-IR** (insulin resistance as measured by the HOMA2 method using fasting glucose [mmol/L] and fasting insulin [pmol/L] for calculation) (Hill et al. 2013)
    - Fasting measurement of glucose and insulin will be the mean of non-missing measurements at -10 and -1 minutes of sMMTT
    - Serum glucose and serum insulin are collected during sMMTT and used for calculation of HOMA2-IR and indices below.
    - If glucose or insulin values are outside of ranges of inputs in HOMA2-IR lookup table, HOMA2-IR will be missing.

  Matsuda Index (postprandial insulin sensitivity from sMMTT) (Schlichtkrull et al. 1965; Matsuda and DeFronzo 1999; Service and O'Brien 2001)

$$ISI - Matsuda = \frac{10{,}000}{\sqrt{G_0 * I_0 * \frac{total\ glucose\ AUC_{0-4hr}}{4hr} * \frac{total\ insulin\ AUC_{0-4hr}}{4hr}}}$$

  where G_0 and I_0 are the glucose concentration (mg/dL) and insulin concentration (μU/mL) at 0 minute (fasting).

  Note: Fasting measurement will be the mean of non-missing measurements at -10 and -1 minutes of sMMTT.

  - **OGIS Index** (Oral Glucose Insulin Sensitivity Index) (Mari et al. 2001)

$$OGIS = \frac{1}{2} \times \left(B + \sqrt{B^2 + 4p_5p_6(G_{120min} - G_{CLAMP})Cl_{OGTT}}\right)$$

  Where

$$B = [p_5(G_{120min} - G_{CLAMP}) + 1] \times Cl_{OGTT}$$

$$Cl_{OGTT} = p4\ \frac{\frac{p_1D_0 - V(G_{180min} - G_{120min})/60min}{G_{120min}} + \frac{p_3}{G_0}}{I_{120min} - I_0 + p_2}$$

  Where $p_1 = 289, p_2 = 270, p_3 = 14{,}000, p_4 = 440, p_5 = 0.000637, p_6 = 117$

  D_0 is an oral glucose dose (expressed in grams per square meter)

  (D_0 = carbohydrate intake [gram] during sMMTT / BSA [m^2], where Gehan and George body surface area formula is BSA=0.1640443958298 × weight[kg]$^{0.515}$ × (height[cm]/100)$^{0.422}$ [Bailey and Briars 1996])

  V=10,000 mL/m^2 (the total glucose distribution volume)

  G_{CLAMP} = 90 mg/dL

G_0 = glucose concentration (mg/dL) at 0 minute (fasting).

I_0 = insulin concentration (µU/mL) at 0 minute (fasting).

G_{120min} = glucose concentration (mg/dL) at 120 min

I_{120min} = insulin concentration (µU/mL) at 120 min

G_{180min} = glucose concentration (mg/dL) at 180 min

Note: Fasting measurement will be the mean of non-missing measurements at -10 and -1 minutes of sMMTT.

- **Stumvoll Index** (Stumvoll et al. 2000; Stumvoll et al. 2001)

$$Stumvoll\ ISI = 0.222 - 0.00333\ BMI - 0.0000779\ I_{120min} - 0.000422\ Age$$

Where I_{120min} = insulin concentration (pmol/L) at 120 min

The above insulin sensitivity parameters during sMMTT will be analyzed using ANCOVA to compare the effect of tirzepatide versus placebo:

- Independent variables: treatment as a factor, baseline as a covariate
- Response variable: change from baseline to Week 18.

Glucose Control

Secondary PD measures to assess the effect of tirzepatide 15 mg versus placebo on parameters of glucose control:

- change from baseline to Week 18 in fasting and postmeal serum glucose during sMMTT
  - fasting glucose
    - mean of nonmissing glucose measurements at -10 and -1 minutes of sMMTT
  - total glucose $AUC_{0\text{-}4hr}$
  - incremental glucose $AUC_{0\text{-}4hr}$

Note: See Section 6.9.1 for AUC calculation details. All these glucose parameters will be reported in the SI Units and conventional units.

The above glucose parameters during sMMTT will be analyzed using ANCOVA to compare the effect of tirzepatide versus placebo:

- Independent variables: treatment as a factor, baseline as a covariate
- Response variable: change from baseline to Week 18.

Time profile plots of serum glucose levels (mean ± SD) during sMMTT (over 11 time points) will be provided by treatment for baseline and Week 18.

- change from baseline through Week 18 in HbA1c

For HbA1c, it is scheduled to be measured at baseline (Day -4) and twice postbaseline (Day 57 [Week 9] and Day 122 [Week 18]). It will be analyzed in a similar manner to secondary PD analysis on body weight using MMRM:

- Independent variables: treatment, time, and treatment-by-time interaction as fixed effects, baseline HbA1c as a covariate, and patient as a random effect
- Response variable: change from baseline in HbA1c

Note: HbA1c will be reported in % and mmol/mol, where

$$\text{HbA1c (mmol/mol)} = 10.93 \times \text{HbA1c (\%)} - 23.5$$

Inferential statistics includes LS means and standard error of the measure for each treatment (tirzepatide, placebo), and the estimated treatment difference (tirzepatide - placebo), corresponding 2-sided 95% CI, and associated p-value.

6.9.2.3. Exploratory Pharmacodynamic Analyses

The retrospective VAS, FCI, and FPQ questionnaires parameters are added below for exploratory analyses:

Retrospective VAS

Exploratory PD measures to assess the effect of tirzepatide 15 mg versus placebo for change from baseline through Week 18 in retrospective VAS (collected 16 times) subscores and overall score:

- subscores:
  - hunger
  - fullness
  - satiety
  - prospective food consumption
- overall score

Note: The retrospective VAS (Womble et al. 2003) will be used to measure the average ratings of appetite that subjects experienced over the past week. The questionnaire consists of only 4 questions, which correspond to 4 appetite subscales (hunger, satiety, fullness, and prospective food consumption). Overall appetite score is calculated as the average of the 4 individual scores (Flint et al. 2000; Flint et al. 2013; van Can et al. 2014):

overall score = (satiety + fullness + [100 - prospective food consumption] + [100 - hunger]) / 4

Summaries of subscores and overall score of retrospective VAS will be provided by treatment for baseline (Day -3), and 15 visits (weekly from Week 1 through Week 15 [Day 1, Day 8, Day 15, Day 22, Day 29, Day 36, Day 43, Day 50, Day 57, Day 64, Day 71, Day 78, Day 85, Day 92, Day 99], and Week 18 [Day 123, Visit 20]), and change from baseline values. Plots of mean scores and mean changes from baseline over time will be provided by treatment.

Food Craving Inventory

Exploratory PD measures to assess the effect of tirzepatide 15 mg versus placebo for change from baseline through Week 18 in FCI 5 subscores and overall score (all are average of nonmissing measurements):

- subscores
  - high fats (8):
    - Fried chicken, Gravy, Sausage, Hot dog, Fried fish, Corn bread, Bacon, Steak
  - sweets (8):
    - Cake, Cinnamon rolls, Ice Cream, Cookies, Chocolate, Donuts, Candy, Brownies
  - carbohydrates/starches (8):
    - Sandwich bread, Rice, Biscuits, Pasta, Pancakes or waffles, Rolls, Cereal, Baked potato
  - fast food fats (4):
    - Pizza, French fries, Hamburger, Chips
  - fruits and vegetables (5):
    - Cooked vegetables, Fruit juices, Raw vegetables, Canned fruit, Raw fruit
- overall score: all 33 items

Note: The 33-item FCI-II (White et al. 2002) is used to measure cravings for specific 5 food groups. It is scaled in a frequency format assessing the frequency of cravings for 33 foods over the past month or since the last time the participant completed the questionnaire. All items are scored in the following 5-point Likert scale: 1= Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Always. Greater scores denote higher levels of craving.

Food Preference Questionnaire

Exploratory PD measures to assess the effect of tirzepatide 15 mg versus placebo for change from baseline through Week 18 in FPQ scores:

- 6 food categories in Table GPGU 6.3:
  - high fat – high sugar, high fat – high complex CHO, high fat – low CHO/high protein, low fat – high sugar, low fat – high complex CHO, low fat – low CHO/high protein)
  - mean rating of food items in each category
- high fat and low fat categories
  - mean rating of food items in each row of Table GPGU 6.3
- high sugar, high complex CHO, and low CHO/high protein categories
  - mean rating of food items in each column of Table GPGU 6.3
- fat preference score (FPS)

$$FPS = \frac{mean\ hedonic\ rating\ for\ high\ fat\ foods}{mean\ hedonic\ rating\ for\ low\ fat\ foods} \times 100$$

  Values of FPS greater than 100 reflect a higher fat preference, and values less than 100 reflect a lower fat preference.

Note: The Food Preference Questionnaire (FPQ) (Geiselman et al. 1998) assesses preferences for 72 foods in a 2 (high-fat, low-fat) by 3 (high-simple sugar, high-complex carbohydrate, low-carbohydrate/high-protein) matrix, with 12 foods in each cell. The macronutrient self-

selection paradigm is an instrument that has to be updated very frequently because of food availability in the grocery stores. During the updates, in the current version of FPQ used in CRF for this study contains 73 food items (where 13 food items in low fat/high sugar category; Table GPGU 6.3).

Subjects rate each food hedonically on a 9-point Likert scale by rating how much they like each food at this moment, with 1 = dislike extremely, 5 = neutral, neither like nor dislike, and 9 = like extremely, and "don't know, never tasted before". "Don't know, never tasted before" are considered missing. The scores for different food categories and FPS (average of nonmissing measurements) will be calculated for statistical analyses.

Table GPGU 6.3 Categorization of FPQ Food Items

	High-Simple Sugar	High-Complex CHO	Low-CHO/High-Protein
High-Fat	Chocolate layer cake Snickers Pecan pie Apple spice cake Vanilla ice cream M&M peanut candies Mounds coconut candy bar Cheesecake, plain Fudge brownie Chocolate cupcake w/ choc icing M&M candies, plain Chocolate ice cream	Pasta with Alfredo Sauce Crescent rolls Cream of celery soup Pizza rolls Onion rings Potato sticks Tortilla chips Fast-food biscuit Stove-Top stuffing Cheese straws French fries Potato salad (mayo type)	American cheese BBQ chicken wings Mozzarella cheese Fried chicken leg Pot roast Hamburger patty Prime rib Sirloin steak Fried egg Peanut butter Fried catfish fillets Scrambled eggs
Low-Fat	Canned pears Canned apricots Banana, fresh Dates, dried Prunes, dried Popsicle, fruit-flavored Cantaloupe, fresh Apple, raw Jelly, any flavor Watermelon, fresh Honeydew melon, fresh Chocolate pudding Vanilla pudding	Cream of wheat Pita bread Long grain rice Leeks Bagel, plain White rice French bread Baked potato, plain Corn, whole kernel Parsnips, cooked Sweet potato, baked, plain Dill pickle	Boiled shrimp Fat-free cheddar cheese Roasted skinless turkey breast Canned shrimp in water Stewed chicken breast Fat-free string cheese Turkey breast canned in water Broiled red snapper Spinach Ground Turkey Boiled crawfish Roasted skinless chicken breast

* Abbreviation: CHO = carbohydrate

The above FCI and FPQ parameters will be analyzed respectively in a manner similar to the secondary PD analysis on body weight using MMRM since FCI and FPQ are scheduled to be measured at baseline (Day -3) and twice postbaseline (Day 50 [Week 8] and Day 123 [Week18]).

- Independent variables: treatment, time, and treatment-by-time interaction as fixed effects, baseline questionnaire parameter as a covariate, and patient as a random effect
- Response variable: change from baseline in questionnaire parameter

Inferential statistics includes LS means and standard error of each of the FCI and FPQ measure for each treatment (tirzepatide, placebo), and the estimated treatment difference (tirzepatide - placebo), corresponding 2-sided 95% CI, and associated p-value.

6.9.2.3.1. Exploratory Pharmacodynamic Analyses on Subcutaneous Adipose Tissue

Exploratory endpoints are change from baseline to Week 18 in exploratory messenger ribonucleic acid (mRNA) expression, lipidomics, metabolomics, or targeted protein assays of subcutaneous adipose tissue biopsy samples, receptor expression, signal transduction, and metabolic changes in adipose tissue. If statistical analyses on these adipose tissue related parameters are conducted, the analyses will be detailed in a separate document and will not require an amendment to this SAP.

6.10. Safety Analyses

Safety measures include, but not limited to, AEs, treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), adverse events of special interest (AESIs), vital signs, and safety laboratory measures.

Unless specified otherwise, safety analyses will be performed on the safety population and presented by treatment group.

Unless specified otherwise, safety listings will display values/events during all study periods. Listings of AEs, death, SAE may include (but not limited to): subject identification (ID) number, age, sex, race, treatment, dose, Medical Dictionary for Regulatory Activities (MedDRA) System Organ Class (SOC), Preferred Term (PT), and time of onset from the first dose of study drug, duration of the AE, seriousness, severity, relatedness to study drug, action taken, and outcome, as appropriate. Additional safety listings will be provided for safety parameters other than AEs in related sections below.

For safety measurements, summary statistics will be provided by treatment. A summary will be provided for AEs with frequency ≥10 patients with such event.

6.10.1. Adverse Events

A listing of adverse events for safety population will be provided, which includes MedDRA PT.

6.10.2. Treatment-Emergent Adverse Events

A TEAE is defined as an AE which first occurs post first dose of study drug or which is present prior to first dose of study drug and becomes more severe post first dose. The maximum severity for each AE during the baseline period including ongoing medical history will be used as baseline severity.

Treatment-emergent adverse events will be summarized by treatment, severity, and relationship with study drug.

6.10.3. Serious Adverse Events

A listing of subjects with SAEs (including death) will be provided.

6.10.4. Adverse Events Leading to Discontinuation

A listing of subjects with AEs leading to discontinuation from study will be provided.

6.10.5. Special Safety Topics

A listing of subjects with all AESIs defined in Section 6.10.5 will be provided.

6.10.5.1. Hypoglycemia

A listing of clinically significant hypoglycemia (plasma glucose <54 mg/dL) and severe hypoglycemia events will be provided. A listing of subjects with hypoglycemic events will be provided. The category of hypoglycemic events (see Table GPGU.6.4 for details) will be presented. A summary will be provided by treatment. The incidence of hypoglycemia will be reported.

Severe/serious hypoglycemia is considered an AESI in this trial.

Table GPGU.6.4. Definitions of Hypoglycemic Event Categories

	Symptoms and/or Signs of Hypoglycemia	Plasma Glucose Level
Glucose alert value	Yes/No/Unknown	≤70 mg/dL (3.9 mmol/L)
Documented symptomatic hypoglycemia	Yes	
Documented asymptomatic hypoglycemia	No	
Documented unspecified hypoglycemia	Unknown	
Clinically significant hypoglycemia	Yes/No/Unknown	<54 mg/dL (3.0 mmol/L)
Documented symptomatic hypoglycemia	Yes	
Documented asymptomatic hypoglycemia	No	
Documented unspecified hypoglycemia	Unknown	

Severe hypoglycemia: defined as an episode with severe cognitive impairment requiring the assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions. Severe hypoglycemia will be reported as an SAE.

Nocturnal hypoglycemia: defined as any hypoglycemic event that occurs between bedtime and waking.

To avoid duplicate reporting, all consecutive glucose values ≤70 mg/dL (3.9 mmol/L) occurring within a 1-hour period may be considered to be a single hypoglycemic event.

6.10.5.2. Pancreatitis

A listing of subjects with pancreatitis (including investigator-reported and adjudicated) will be provided. Adjudication assessment results will be reported in the listing.

Treatment-emergent adjudicated-confirmed pancreatitis will be considered as AESI.

6.10.5.3. Thyroid Malignancies and C-Cell Hyperplasia

A listing of subjects with thyroid malignancies and C-cell hyperplasia (search criteria provided in Appendix 1) will be provided.

Thyroid malignancies and C-cell hyperplasia will be considered as AESIs.

6.10.5.4. Major Adverse Cardiovascular Events

A listing of subjects with MACE (including investigator-reported and adjudicated) will be provided. Adjudication assessment results will be reported in the listing.

Only positively adjudicated MACE will be considered as an AESI.

6.10.5.5. Arrhythmias and Cardiac Conduction Disorders

A listing of subjects with arrhythmias and cardiac conduction disorders (search criteria provided in Appendix 1) will be provided.

Severe/serious treatment-emergent arrhythmias and cardiac conduction disorders will be considered as AESI. The cardiovascular events will include clinically relevant rhythm and conduction disorders.

6.10.5.6. Hypersensitivity Events

A listing of subjects with hypersensitivity reactions (search criteria provided in Appendix 1) will be provided.

Only serious/severe cases of hypersensitivity will be considered as AESIs.

6.10.5.7. Injection-Site Reactions

Injection-site assessments for local tolerability will be conducted, when reported as an AE from a patient, or a clinical observation from an investigator.

A listing of subjects with reported injection-site reactions (eg, edema, erythema, induration, itching, and pain) will be provided. Detailed search criteria can be found in Appendix 1.

Only the severe/serious injection site reactions (eg, abscess, cellulitis, erythema, hematomas/hemorrhage, exfoliation/necrosis, pain, subcutaneous nodules, swelling, indurating, inflammation) will be considered as AESIs.

6.10.5.8. Hepatobiliary Disorders

A listing of subjects with events of biliary colic, cholecystitis, or other suspected events related to gallbladder disease will be provided. Detailed search criteria can be found in Appendix 1.

Severe/serious hepatobiliary disorders will be considered as AESIs.

6.10.5.8.1. Hepatic Monitoring

The subjects' liver disease history and associated person liver disease history data will be listed. Concomitant medications acetaminophen/paracetamol, which have potential to cause hepatotoxicity, will be listed. Results from any hepatic monitoring procedures, such as a

magnetic resonance elastography (MRE) scan, and biopsy assessments will be listed, if performed.

Hepatic risk factor assessment data will be listed. Liver related signs and symptoms data will be listed, and it will be summarized by treatment if ≥10 subjects with such data. Alcohol and recreational drug use data may be listed.

All hepatic chemistry, hematology, coagulation, and serology data will be listed. Values outside the reference ranges will be flagged on the individual subject data listings.

6.10.5.9. Severe/Serious Gastrointestinal Adverse Events

A listing of subjects with severe/serious gastrointestinal AEs, such as nausea, vomiting, and diarrhea will be provided.

Only the PTs in the gastrointestinal MedDRA SOC with serious/severe cases will be considered as AESIs.

6.10.5.10. Acute Renal Events

A listing of subjects with acute renal events (search criteria provided in Appendix 1) will be provided.

Severe/serious acute renal events will be considered as AESI.

6.10.5.11. Major Depressive Disorder/Suicidal Ideation Monitoring Using the C-SSRS and PHQ-9 Questionnaires

Only *serious* AEs elicited through the Columbia-Suicide Severity Rating Scale (C-SSRS), or Patient Health Questionnaire (PHQ-9) are to be recorded as AEs via the eCRF and reported to Lilly or its designee within 24 hours as SAEs.

The serious/severe major depressive disorder/suicidal ideation or behavior will be considered as AESI. Adverse events will be searched using MedDRA PT terms. Detailed searching criteria can be found in Appendix 1.

6.10.5.11.1. Patient Health Questionnaire

The PHQ-9 is a validated self-report screening tool that assesses the presence and intensity of depressive symptoms. The PHQ-9 which includes 9 questions (corresponding to 9 Diagnostic and Statistical Manual-IV depression criteria) was developed for use in primary care settings (Kroenke et al. 2001).

These items are scored in the following 4-point Likert scale: 0 = Not at all, 1 = Several days, 2 = More than half the days, and 3 = Nearly every day. Total score for the PHQ-9 ranges from 0 to 27. Greater scores denote higher level of depression.

For each subject, total scores of nonmissing items in PHQ-9 are calculated and categorized as

- none (not depressed): 0 through 4
- mild depression: 5 through 9
- moderate depression: 10 through 14

- moderately severe depression: 15 through 19
- severe depression: 20 through 27

Note: The last question "If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?" is not included as one of these 9 questions to calculate total score.

PHQ-9 is administered at screening (baseline), Day 50 (Week 8, Visit 10), Day 106 (Week 16, Visit 18), and follow-up/early termination visit. A listing of patients with treatment-emergent depression (ie any increase in PHQ-9 categories) will be provided. The count and percentage of patients within each above category will be summarized by treatment and by visit.

6.10.5.11.2. Suicidal Ideation and Behavior Solicited Through C-SSRS

C-SSRS is administered at screening, Day -5 (lead-in, baseline), Day 22 (Week 4, Visit 6), Day 50 (Week 8, Visit 10), Day 78 (Week 12, Visit 14), Day 106 (Week 16, Visit 18), and follow-up/early termination visit.

The suicidal ideation, suicidal behavior, and non-suicidal self-injurious behavior are determined*:

- **Suicidal ideation (5):** A "yes" answer at any time during study to any 1 of the 5 suicidal ideation questions (1 "wish to dead", 2 "non-specific active suicidal thoughts", 3 "active suicidal ideation with any methods (not plan) without intent to act", 4 "active suicidal ideation with some intent to act, without specific plan", 5 "active suicidal ideation with specific plan and intent") on the C-SSRS.
- **Suicidal behavior (6):** A "yes" answer at any time during study to any 1 of the 5 suicidal behavior questions (6 "actual attempt", 8 "interrupted attempt", 9 "aborted attempt", 10 "preparatory acts or behavior", 11 "suicidal behavior" 12 "completed suicide") on the C-SSRS.
- **Non-suicidal self-injurious behavior (1):** A "yes" answer to 7 "has subject engaged in Non-Suicidal Self-Injurious behavior?"

Note: At screening and lead in visit (Day -5), C-SSRS baseline/screening version will be used, and the ideation and behavior will be assessed per subject answers on "past 1 month." At postbaseline visits, C-SSRS since last visit version will be used, and the ideation and behavior will be assessed per subject answers on "since last visit."

Baseline C-SSRS is administered at lead in visit (Day -5).

A listing of subjects with any treatment-emergent suicidal ideation, suicidal behavior, or non-suicidal self-injurious behavior during the study based on C-SSRS will be provided. The category of suicidal ideation, suicidal behavior, or non-suicidal self-injurious behavior is reported in listing. Data from C-SSRS of all visits are displayed for these subjects.

If any data were collected from self-harm supplemental and self-harm follow-up questionnaires (single CRF) which are triggered by any self-harm through C-SSRS, it will be listed.

6.10.6. Vital Signs

Summaries of vital signs will be provided by treatment for actual values of baseline (Day -2), postbaseline values at Day 8 (Week 2), Day 15 (Week 3), Day 22 (Week 4), Day 29 (Week 5), Day 43 (Week 7), Day 57 (Week 9), Day 85 (Week 13), Day 113 (Week 17), Day 123 (Week 18), Day 124 (Week 18), Day 125 (Week 18), follow-up visit and change from baseline values. Plots of mean vital signs and mean changes from baseline over time will be provided by treatment.

The treatment-emergent abnormal vital signs will be listed. The criteria for identifying patients with treatment-emergent vital sign abnormalities are stated in Table GPGU.6.5.

Table GPGU.6.5. Categorical Criteria for Treatment-Emergent Abnormal Blood Pressure and Pulse Measurements

Parameter	Low	High
Systolic BP (mmHg) (Supine or sitting – forearm at heart level)	≤90 and decrease from baseline ≥20	≥140 and increase from baseline ≥20
Diastolic BP (mmHg) (Supine or sitting – forearm at heart level)	≤50 and decrease from baseline ≥10	≥90 and increase from baseline ≥10
Pulse (bpm) (Supine or sitting)	<50 and decrease from baseline ≥15	>100 and increase from baseline ≥15

Abbreviations: BP = blood pressure; bpm = beats per minute.

6.10.7. Electrocardiogram

Electrocardiograms (ECGs) will be performed for safety monitoring purposes only. Any clinically significant ECG findings entered as AEs will be included in AE summaries.

6.10.8. Safety Laboratory Parameters

All laboratory data will be reported in the International System of Units and conventional units.

Descriptive summaries of clinical chemistry, hematology, and endocrine (calcitonin) data and their changes from baseline will be provided at each visit by treatment.

Additionally, clinical chemistry, hematology, urinalysis, and endocrine (calcitonin) data outside the reference ranges will be listed.

If any safety lab measurements are (1) below the quantification limit (eg <QL), ½ × QL may be used for the calculation of summary statistics; (2) above the quantification limit (eg >QL), 1.1 × QL may be used for the calculation of summary statistics, if deemed appropriate.

6.11. Evaluation of Immunogenicity

Baseline immunogenicity sample is collected on Day 1 pre-dose.

For TE-ADA+ subjects, the distribution of maximum titers may be described.

The frequency and percentage of patients with preexisting ADA and with TE-ADA+ to tirzepatide may be tabulated.

If cross-reactive antibodies to native GIP and GLP-1, or neutralizing antibodies against native GLP-1 and GIP assays are performed, the frequency and percentage of each may be tabulated.

A listing will be provided of immunogenicity assessments. This includes the tirzepatide concentration from a simultaneous pharmacokinetic sample and the clinical interpretation result (ADA Present, ADA Not Present, ADA Inconclusive, or Missing). A listing of TEAE for patients with TE ADA+ or Injection Site Reaction or Potential Hypersensitivity may be provided.

Cases of TE-ADA that are associated with AEs of either severe/serious hypersensitivity or severe/serious injection site reaction will be classified as AESI.

6.12. Interim Analyses

No interim analyses are planned for this study. If an unplanned interim analysis is deemed necessary, the Lilly clinical pharmacologist, clinical research scientist/clinical research physician, investigator, or designee will consult with the appropriate medical director or designee to determine if it is necessary to amend the protocol.

6.13. COVID-19 Impact Assessment

A listing of subjects with adverse events related to COVID-19, including death due to COVID-19, serious COVID-19 adverse events, and COVID-19 adverse events, may be provided. Protocol deviations that study disruptions related to COVID-19 may be described in CSR.

7. Unblinding Plan

The blinding/unblinding plan is not part of this SAP. The approved blinding/unblinding plan is stored in Leo.

8. Changes from the Protocol Specified Statistical Analyses

Changes from the protocol specified statistical analyses are as follows:

- In protocol Section 1 and Section 10.3.3.2, it states "The primary analysis will be performed in all randomized subjects with evaluable data (modified intent-to-treat [mITT] population)". It is clarified in this SAP that PD analyses will be conducted on the PD analysis set (ie evaluable data from all randomized subjects who are exposed at least 1 dose of study drug). This is consistent with protocol Section 10.3 "Pharmacodynamic analyses will be conducted on data from all subjects who receive at least 1 dose of the IP and have evaluable data."
- In protocol Section 10.1 Sample Size Determination, it is updated to "Up to 56 subjects are planned to be randomized so that approximately 46 subjects complete the study assuming 15% discontinuation rate. These 56 subjects will be randomized to QW tirzepatide or placebo in a 1:1 ratio. The estimated standard deviation of the change in SMR from baseline is 117 kcal/day (Heilbronn et al. 2006). With an expected treatment difference of 100 kcal/day, this provides at least 80% power for the comparison of tirzepatide versus placebo based on 2 sample t-test using 2-sided test at alpha level of 0.05 for the assessment of the primary endpoint."
- In protocol appendix Section 6.4, it states "The Food Preference Questionnaires (FPQ) (Geiselman et al. 1998) assesses preferences for 72 foods in a 2 (high-fat, low-fat) by 3 (high-simple sugar, high-complex carbohydrate, low-carbohydrate/high-protein) matrix". It is updated and clarified in this SAP "The macronutrient self-selection paradigm (MSSP) is an instrument that has to be updated very frequently because of food availability in the grocery stores. During the updates, in the current version of FPQ used in CRF for this study contains 73 food items (where 13 food items in low fat/high sugar category)."
- In protocol Section 7.6 Treatment Compliance, it states "Subjects will be considered compliant if they received at least 75% of their scheduled doses for each dosing interval (Weeks 1 to 2, Weeks 3 to 4, Weeks 5 to 8, and Weeks 9 to endpoint) during the treatment period." This is updated in SAP that subjects will be considered treatment compliant if taking ≥75% of their scheduled doses for 18-week treatment period taking into account early termination (specifically excluding the period after early termination), for statistical summaries.

9. References

Bailey BJ, Briars GL. Estimating the surface area of the human body. *Stat Med.* 1996;15(13):1325-1332. https://doi.org/10.1002/(SICI)1097-0258(19960715)15:13%3C1325::AID-SIM233%3E3.0.CO;2-K

Flint A, Raben A, Blundell JE, et al. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *Int J Obes Relat Metab Disord.* 2000;24(1):38-48. https://doi.org/10.1038/sj.ijo.0801083

Flint A, Kapitza C, Zdravkovic M. The once-daily human GLP-1 analogue liraglutide impacts appetite and energy intake in patients with type 2 diabetes after short-term treatment. *Diabetes Obes Metab*. 2013;15:958-962. https://doi.org/10.1111/dom.12108

Frayn, KN. Calculation of substrate oxidation rates in vivo from gaseous exchange. *J Appl Physiol Respir Environ Exerc Physiol*. 1983;55(2):628-634. https://doi.org/10.1152/jappl.1983.55.2.628

Geiselman PJ, Anderson AM, Dowdy ML, et al. Reliability and validity of a macronutrient self-selection paradigm and a food preference questionnaire. *Physiol Behav*. 1998;63(5):919-928. https://doi.org/10.1016/S0031-9384(97)00542-8

Heilbronn LK, de Jonge L, Frisard MI, et al. Effect of 6-month calorie restriction on biomarkers of longevity, metabolic adaptation, and oxidative stress in overweight individuals: a randomized controlled trial. *JAMA*. 2006;295(13):1539-1548. https://doi.org/10.1001/jama.295.13.1539

Hill NR, Levy JC, Matthews DR. Expansion of the homeostasis model assessment of β-cell function and insulin resistance to enable clinical trial outcome modeling through the interactive adjustment of physiology and treatment effects: iHOMA2. *Diabetes Care*. 2013;36(8):2324-2330. https://doi.org/10.2337/dc12-0607

Jéquier E, Acheson K, Schutz Y. Assessment of energy expenditure and fuel utilization in man. *Annu Rev Nutr*. 1987;7:187-208. https://doi.org/10.1146/annurev.nu.07.070187.001155

Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Mari A, Pacini G, Murphy E, Ludvik B, Nolan J. A model-based method for assessing insulin sensitivity from the oral glucose tolerance test. *Diabetes Care*. 2001;24:539-548. https://doi.org/10.2337/diacare.24.3.539

Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care*. 1999;22(9):1462-1470. https://doi.org/10.2337/diacare.22.9.1462

Schlichtkrull J, Munck O, Jersild M. The M-value, an index of blood-sugar control in diabetics. *Acta Med Scand*. 1965;177:95-102. https://doi.org/10.1111/j.0954-6820.1965.tb01810.x

Service FJ, O'Brien PC. The relation of glycaemia to the risk of development and progression of retinopathy in the diabetic control and complications trial. *Diabetologia*. 2001;44(10):1215-1220. https://doi.org/10.1007/s001250100635

Stumvoll M, Mitrakou A, Pimenta W, Jenssen T, Yki-Jarvinen H, Van Haeften T, Renn W, Gerich J. Use of the oral glucose tolerance https://doi.org/10.2337/diacare.23.3.295 test to assess insulin release and insulin sensitivity. *Diabetes Care*. 2000;23(3):295-301. https://doi.org/10.2337/diacare.23.3.295

Stumvoll M, Van Haeften T, Fritsche A, Gerich J. Oral glucose tolerance test indexes for insulin sensitivity and secretion based on various availabilities of sampling times. *Diabetes Care*. 2001;24(4):796-797. https://doi.org/10.2337/diacare.24.4.796

van Can J, Sloth B, Jensen CB, et al. Effects of the once-daily GLP-1 analog liraglutide on gastric emptying, glycemic parameters, appetite and energy metabolism in obese, non-diabetic adults. *Int J Obes (Lond)*. 2014;38(6):784-793. https://doi.org/10.1038/ijo.2013.162

Weir JB. New methods of calculating metabolic rate with special reference to protein metabolism. *J Physiol*. 1949;109(1-2):1-9. https://doi.org/10.1113/jphysiol.1949.sp004363

White MA, Whisenhunt BL, Williamson DA, et al. Development and validation of the food-craving inventory. *Obes Res*. 2002;10(2):107-114. https://doi.org/10.1038/oby.2002.17

Womble LG, Wadden TA, Chandler JM, et al. Agreement between weekly vs. daily assessment of appetite. *Appetite*. 2003;40(2):131-135. https://doi.org/10.1016/s0195-6663(02)00170-8

10. Appendices

Appendix 1. Search Criteria For Special Safety Topics

The search criteria for AEs of special safety topics and AESIs are stored in CLUWE:\\statsclstr\lillyce\qa\ly3298176\common\AESI_Lab\Search criteria AESIs_TZP.xlsx

Leo Document ID = 4ef305f6-7679-449b-8772-b24308589e24

Approver: PPD

Approval Date & Time: 14-Jan-2022 20:53:05 GMT

Signature meaning: Approved