I8F-MC-GPHH Statistical Analysis Plan Version 3

# Effect of Tirzepatide on Energy Intake and Appetite-and Reward-Related Brain Areas in Overweight/Obese Subjects: A Placebo-Controlled 6-Week Study With Functional MRI

NCT04311411

Approval Date: 20-Apr-2023

# 1. Statistical Analysis Plan
# I8F-MC-GPHH: Effect of Tirzepatide on Energy Intake and Appetite- and Reward-Related Brain Areas in Overweight/Obese Subjects: A Placebo-Controlled 6-Week Study with Functional MRI



LY3298176 for Overweight/Obesity

Phase 1, multiple-center, randomized, partially blinded, placebo-controlled, parallel-arm study, with a positive control, liraglutide, in overweight/obese subjects to compare tirzepatide to placebo.


Eli Lilly and Company
Indianapolis, Indiana USA 46285


Protocol I8F-MC-GPHH
Phase 1

**Document ID:** VV-CLIN-109361



Approved on 20 Apr 2023 GMT

# 2. Table of Contents

**Section** **Page**

## Table of Contents

**Tables** **Page**

## Table of Contents

**Figure** **Page**

# 3. List of Abbreviations

| Term | Definition |
|---|---|
| **ADA** | anti-drug antibody |
| **BIS** | Barratt Impulsiveness Scale |
| **BMI** | body mass index |
| **BOLD fMRI** | blood-oxygenation-level-dependent functional magnetic resonance imaging |
| **AE** | adverse event |
| **AESI** | adverse event of special interest |
| **CI** | confidence interval |
| **COVID-19** | coronavirus disease 2019 |
| **ECG** | electrocardiogram |
| **FCI** | Food Craving Inventory |
| **FCQ-S** | Food Craving Questionnaire-State |
| **GIP** | glucose-dependent insulinotropic polypeptide |
| **GLP-1** | glucagon-like peptide 1 |
| **IPD** | important protocol deviation |
| **Lilly** | Eli Lilly and Company |
| **LSM** | least squares means |
| **MedDRA** | Medical Dictionary for Regulatory Activities |
| **MMRM** | mixed effects model of repeated measures |
| **MRI** | magnetic resonance imaging |
| **PD** | pharmacodynamic(s) |
| **PK** | pharmacokinetic(s) |
| **PT** | Preferred Term |
| **PYY** | peptide YY |
| **QD** | once daily |
| **QL** | quantification limit |

| Term | Definition |
|---|---|
| **QW** | once weekly |
| **ROI** | region of interest |
| **SAE** | serious adverse event |
| **SAP** | statistical analysis plan |
| **SOC** | System Organ Class |
| **TE-ADA(+)** | treatment-emergent anti-drug antibody (positive) |
| **TEAE** | treatment-emergent adverse event |
| **TFEQ** | Three-Factor Eating Questionnaire |
| **VAS** | visual analog scale |

# 4. Revision History

Statistical analysis plan version 1 was approved on 25 June 2020 prior to the first patient visit.

Statistical analysis plan version 2 was approved on 09 June 2022 prior to database lock.

The major revisions in version 2 include

- modified study procedure and schedule of activities based on Protocol (f)
- updated sample size to approximately 111 subjects (37 subjects per treatment arm) planned to be randomly assigned, and
- reorganized and rephrased the PD analyses section.

Statistical analysis plan version 3 was approved prior to database lock.

Major revisions in version 3 include

- Change of specification of the BOLD fMRI analysis variables. Contrasts between conditions are defined as the mean difference as opposed to mean percent difference. This is because dividing by the non-food contrast was causing falsely large numbers in scans where the subject had low activation to non-food objects (that is, division by number close to 0).

# 5. Study Objectives

Table GPHH.5.1 shows the objectives and endpoints of the study.

**Table GPHH.5.1. Objectives and Endpoints**

| Objectives | Endpoints |
|---|---|
| **Primary**<br>To compare the effect of 5-mg tirzepatide versus placebo, at Week 3, on energy intake in a clinical setting | • Change from baseline in energy intake (kcal) as assessed by ad-libitum food-intake test |
| **Secondary**<br>To compare the effect of 5-mg tirzepatide versus placebo, at Week 3 on | |
| • Parameters of central reward and appetite circuits in the fasting states using BOLD fMRI | • Change from baseline in BOLD fMRI activation to images of highly palatable food (high fat-high sugar and high fat-high carbohydrate) during the fasting state in the brain reward areas (insula, medial frontal gyrus, superior temporal gyrus, precentral gyrus, and cingulate gyrus) |
| • Parameters of behavioral appetite assessments | • Change from baseline in fasting and postprandial appetite VAS, and fasting FCI, FCQ-S, Eating Inventory, and Power of Food Scale questionnaires |
| **Exploratory**<br>To compare the effect of 5- or 10-mg tirzepatide versus placebo at Week 3 and/or Week 6 on | |
| • Parameters of central reward and appetite circuits in the fasting state using BOLD fMRI (Week 6/10 mg) | • Change from baseline in BOLD fMRI activation to images of highly palatable food (high fat-high sugar and high fat-high carbohydrate) versus non-food objects during the fasting state in the brain reward areas (insula, medial frontal gyrus, superior temporal gyrus, precentral gyrus, cingulate gyrus) |
| • Parameters of behavioral appetite assessments (Week 6/10 mg) | • Change from baseline in fasting and postprandial appetite VAS, and fasting FCI, FCQ-S, Eating Inventory, and Power of Food Scale questionnaires |
| • Energy intake in a clinical setting (Week 6/10 mg) | • Change from baseline in energy intake (kcal) as assessed by ad-libitum food-intake test |

| Objectives | Endpoints |
|---|---|
| • Parameters of central reward and appetite circuits in the fasting state using BOLD fMRI (Week 3/5 mg and Week 6/10 mg) | • Change from baseline in BOLD fMRI activation to images of highly palatable food (high fat-high sugar and high fat-high carbohydrate) during fasting state in the hippocampus, putamen, orbitofrontal cortex, and ventral striatum brain regions |
| • Impulsivity (Week 3/5 mg and Week 6/10 mg) | • Change from baseline in BIS questionnaire |
| • Appetite and metabolism regulating hormones (Week 3/5 mg and Week 6/10 mg) | • Change from baseline in ghrelin, glucagon, glucose, and triglycerides (fasting), insulin (fasting and postprandial), and amylin, GIP, GLP-1, leptin, pancreatic polypeptide, and PYY (postprandial) |
| To compare the effect of 5- and 10-mg tirzepatide versus liraglutide and tirzepatide versus placebo at Week 3 and/or Week 6 on | |
| • Energy intake in a clinical setting (Week 3/5 mg and Week 6/10 mg) | • Changes from baseline in energy intake (kcal) as assessed by ad-libitum food-intake test |
| • Parameters of behavioral appetite assessments (Week 3/5 mg and Week 6/10 mg) | • Change from baseline in fasting and postprandial appetite VAS, and the fasting FCI, FCQ-S, Eating Inventory, and Power of Food Scale questionnaires |
| • Parameters of central reward and appetite circuits in the fasting states using BOLD fMRI (Week 3/5 mg and Week 6/10 mg) | • Change from baseline in BOLD fMRI activation to images of highly palatable food (high fat-high sugar and high fat-high carbohydrate during the fasting state in the brain reward areas [insula, medial frontal gyrus, superior temporal gyrus, precentral gyrus, cingulate gyrus, hippocampus, putamen, orbitofrontal cortex, and ventral striatum]) |
| • Appetite and metabolism regulating hormones (Week 3/5 mg and Week 6/10 mg) | • Change from baseline in ghrelin, glucagon, glucose, and triglycerides (fasting), insulin (fasting and postprandial), and amylin, GIP, GLP-1, leptin, pancreatic polypeptide, and PYY (postprandial) |

| Objectives | Endpoints |
|---|---|
| To assess the safety and tolerability of tirzepatide. | • Adverse events<br>• Safety laboratory parameters<br>• Frequency of treatment-emergent anti-tirzepatide antibodies |

Abbreviations: BIS = Barratt Impulsiveness Scale; BOLD = blood-oxygenation-level-dependent; FCI = Food Craving Inventory; FCQ-S = Food Craving Questionnaire-State; fMRI = functional MRI; GIP = glucose-dependent insulinotropic polypeptide; GLP-1 = glucagon-like peptide 1; MRI = magnetic resonance imaging; PYY = peptide YY; VAS = visual analog scale.

# 6. Study Design

## 6.1. Study Design and Treatment

This is a Phase 1, multicenter, randomized, partially blinded, placebo-controlled, parallel-arm study, with a positive control, liraglutide, in overweight/obese subjects. The primary objective of this study is to compare the effect of tirzepatide versus placebo on energy intake, as assessed by an ad-libitum food-intake test, at Week 3 of treatment at a therapeutic dose of 5 mg. Secondary objectives will compare these treatment groups for parameters of central reward, appetite circuits, and behavioral appetite assessments at Week 3.

Additional exploratory objectives will assess the effect of 10-mg tirzepatide versus placebo on energy intake, parameters of central reward and appetite, and behavioral appetite assessments at Week 6. Further exploratory analysis will assess the effect of tirzepatide versus liraglutide and liraglutide versus placebo on energy intake, parameters of central reward and appetite signaling, and behavioral appetite assessments at Week 3 and/or 6.

The study will consist of the following periods: approximately 5-week screening period, 5-day lead-in period, 6-week treatment period, and a 4-week safety follow-up period. Subjects will be randomized in a 1:1:1 ratio to placebo, tirzepatide, or liraglutide. Investigators and subjects will be blinded to tirzepatide and placebo treatment; however, liraglutide treatment will be open label, so the study is considered partially blinded. The randomization will be stratified by baseline BMI (27 to less than 30 $kg/m^2$, 30 to less than 35 $kg/m^2$, and 35 to 50 $kg/m^2$).

Tirzepatide QW dosing will start at 5 mg for 3 weeks followed by a dose escalation to 10 mg for 3 weeks. Subjects will return to the clinical research unit each week for the QW administration of tirzepatide or placebo.

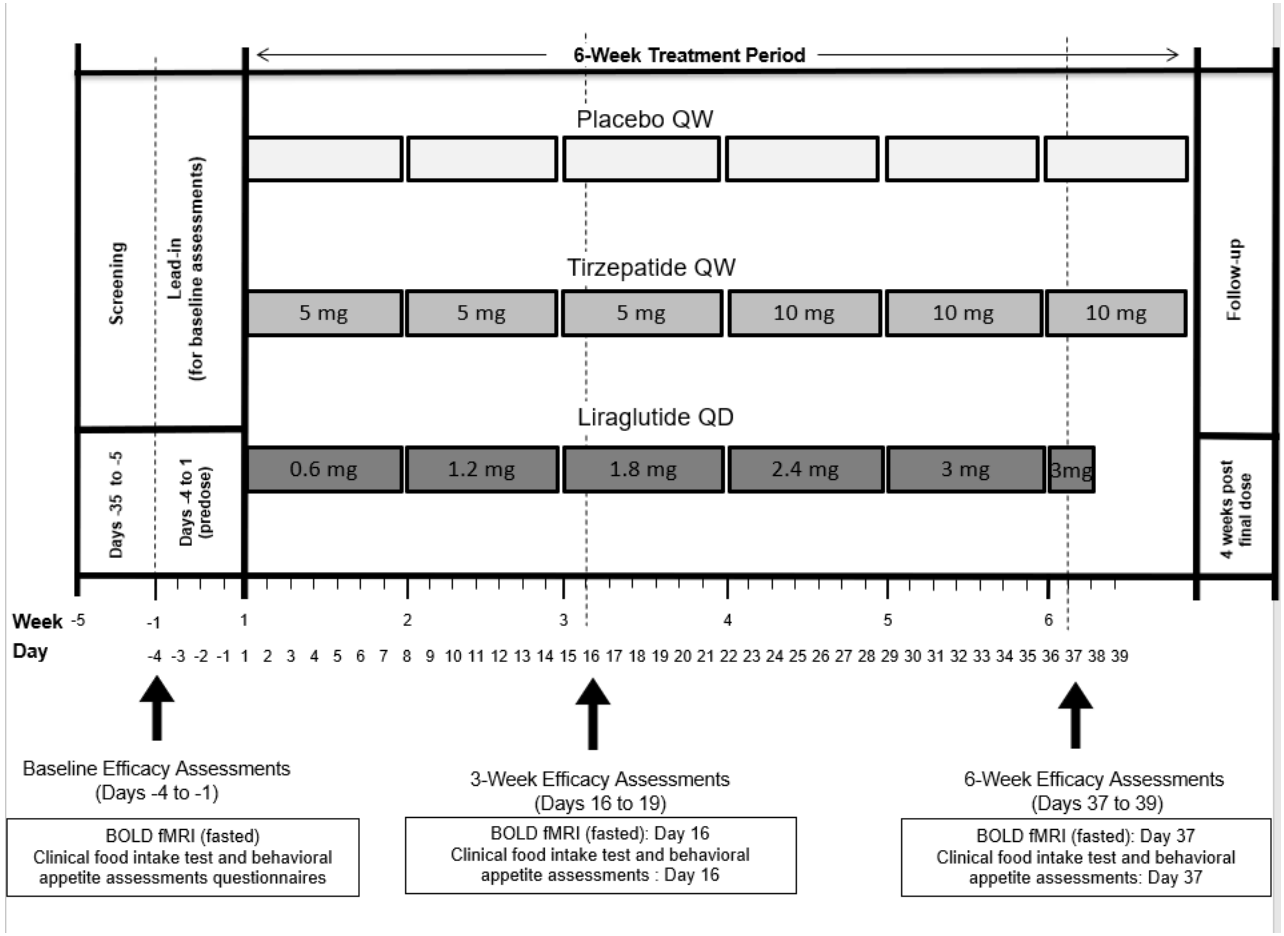
Liraglutide QD dosing will start at 0.6 mg for 1 week followed by weekly, step-wise dose escalations to 1.2, 1.8, and 2.4 mg, until a 3-mg QD dose is reached. The 3-mg dose of liraglutide will be maintained for 10 days. Subjects will self-administer liraglutide at home in the evening.

Figure GPHH.6.1 illustrates the study design.

## 6.2. Determination of Sample Size

Up to 111 subjects (37 subjects per treatment arm) are planned to be randomized so that approximately 93 subjects (31 subjects per treatment arm) complete the study, assuming a 15% discontinuation rate.

These 111 subjects will be randomized to tirzepatide, liraglutide, or placebo in a 1:1:1 ratio. The estimated variability (standard deviation) of the change in energy intake from baseline is 289 kcal on the results from GPGC and blinded data from study GPGT Study I8F-MC-based based on an expected treatment difference of 212 kcal, this provides at least 80% power for the comparison of tirzepatide versus placebo based on a 2-sample t-test using a 2-sided test at an $\alpha$ level of 0.05.




Abbreviations: BOLD = blood-oxygenation-level-dependent; fMRI = functional MRI; MRI = magnetic resonance imaging; QD = once daily; QW = once weekly.

**Figure GPHH.6.1. Illustration of study design for Protocol I8F-MC-GPHH.**

# 7. A Priori Statistical Methods

## 7.1. Populations for Analyses

For the purpose of analysis, Table GPHH.7.1 defines 3 analysis sets.

**Table GPHH.7.1. Analysis Populations and Analysis Datasets**

| Population/Data Set | Description |
|---|---|
| Randomized population | All subjects who are randomly assigned to a treatment arm |
| Safety population | All randomized subjects who receive at least 1 dose of the randomly assigned IP, regardless of whether they completed all protocol requirements |
| Pharmacodynamic analysis set | All randomized subjects who receive at least 1 dose of the randomly assigned IP and have evaluable data |

Abbreviations: IP = investigational product; MRI = magnetic resonance imaging; PD = pharmacodynamic.
Note: Protocol deviations will be considered for their severity/impact and taken into consideration if subjects should be excluded from PD analysis set for MRI data-related analyses.

## 7.2. General Considerations

Statistical analysis of this study will be the responsibility of 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 this SAP.

Unless otherwise specified, safety analyses will be conducted on the safety population (Table GPHH.7.1), and PD analyses will be conducted on the PD analysis set (Table GPHH.7.1).

Unless otherwise specified, the baseline values used for analyses will be the last scheduled, nonmissing value obtained for each subject before first dose.

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

## 7.3. Study Participant Characteristics

Subject baseline demographic and clinical characteristics will be summarized for the randomized population by study treatment (including overall). Specifically, the following baseline parameters will be included in (but not limited to) the summary report:

- age (years)
- gender (male and female)
- race
- ethnicity (Hispanic and Non-Hispanic)
- body weight (kg)
- height (cm)
- BMI ($kg/m^2$)
- BMI (27 to less than 30, 30 to less than 35, and 35 to 50 $kg/m^2$), and

- waist circumference (cm).

## 7.4. Study Participant Disposition

A listing and a summary of subject disposition for the randomized population (that is, subjects who have been assigned to a treatment) by actual study treatment will be provided.

A listing of patients 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, race, treatment, and study period.

## 7.5. Concomitant Therapy

Concomitant medications will be listed and summarized by treatment and World Health Organization Drug Dictionary Medication Class and PT for the safety population. In addition, concomitant medication taken by 10 or more subjects, regardless of treatment, will also be summarized.

## 7.6. Treatment Exposure and Compliance

### *7.6.1. Treatment Exposure*

A listing of enrolled subjects who had inadvertently received the incorrect study treatment anytime during the study will be provided, if applicable.

The duration of exposure to study treatment (tirzepatide, liraglutide, 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 treatment will be summarized by treatment using the safety population.

### *7.6.2. Treatment Compliance*

Percent compliance will be listed for each patient and summarized by treatment on the safety population. Percent compliance and treatment compliance will be summarized by treatment for each period and by treatment only.

Percent compliance will be calculated for each subject 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% or more of their scheduled doses for the 6-week treatment period, considering early termination (specifically excluding the period after early termination).

## 7.7. Important Protocol Deviations

Important protocol deviations are defined as deviations from the study protocol that may significantly compromise the data integrity and/or patient safety. The details of identification of IPDs are provided in a separate document (that is, Risk and Issue Mitigation and Management Tool [RIMM]). A listing/table of IPDs will be provided by the study manger after database lock.

## 7.8. Pharmacokinetic Analyses

### *7.8.1. Pharmacokinetic Parameter Estimation*

Sparse PK samples will be collected across the 6-week treatment period. Tirzepatide concentrations will be determined to support an understanding of tirzepatide exposure over the treatment duration and compare with expected tirzepatide PK.

### *7.8.2. Pharmacokinetic Statistical Inference*

No summaries or analyses of PK parameters are planned. Tirzepatide concentrations may be summarized by week.

## 7.9. Pharmacodynamic Analyses

### *7.9.1. Pharmacodynamic Parameter Estimation*

Pharmacodynamic parameters will be measured at lead-in (baseline), Week 3, and/or Week 6, to assess the effect of study treatments on measures of appetite and food intake, food intake-related and impulsiveness questionnaires, central reward and appetite circuits in the brain, and hormones and metabolites related to appetite regulation. The change from baseline to Week 3 and Week 6 (if applicable) will be calculated for all continuous PD parameters.

**Table GPHH.7.2. Study Period and PD Measurements**

| | Period | Visit (Study Day) | Week | PD Measurements |
|---|---|---|---|---|
| 1 | Lead-in (baseline) | 2 (between Day -4 and Day -1) | -1 | • Ad-libitum food intake test at lunch (clinic-based) with appetite VAS<br>• FCI, FCQ-S, Eating Inventory, BIS and Power of Food Scale questionnaires<br>• Retrospective appetite VAS<br>• BOLD fMRI<br>• Hormones and metabolites for appetite and metabolism regulation in the fasting and postprandial states |
| 2 | Treatment | 6 (16 ±1) | Week 3 | • Ad-libitum food intake test at lunch (clinic-based) with appetite VAS<br>• FCI, FCQ-S, Eating Inventory, BIS and Power of Food Scale questionnaires<br>• Retrospective appetite VAS |

| | Period | Visit (Study Day) | Week | PD Measurements |
|---|---|---|---|---|
| | | | | • BOLD fMRI<br>• Hormones and metabolites for appetite and metabolism regulation in the fasting and postprandial states |
| 3 | Treatment | 11 (37 ±1) | Week 6 | • Ad-libitum food intake test at lunch (clinic-based) with appetite VAS<br>• FCI, FCQ-S, Eating Inventory, BIS and Power of Food Scale questionnaires<br>• Retrospective appetite VAS<br>• BOLD fMRI<br>• Hormones and metabolites for appetite and metabolism regulation in the fasting and postprandial states |
| 4 | Follow-up | 12 (59 ±1 if TZP or placebo and 61 ±1 if liraglutide) | Follow-up Week 3 | • Ad-libitum food intake test at lunch (clinic-based) |
| 5 | Follow-up | All days between 12 and 13 (~4 days) | | • Ad-libitum food intake assessment (free-living) |
| 6 | Follow-up | 13 (64 ±1 if TZP/ or placebo and 66 ±1 if liraglutide) | Follow-up Week 4 | • Ad-libitum food intake test at lunch (clinic-based) |

Abbreviations: ~ = approximately; BIS = Barratt Impulsiveness Scale; BOLD = blood-oxygenation-level-dependent; FCI = Food Craving Inventory; FCQ-S = Food Craving Questionnaire-State; fMRI = functional MRI; MRI = magnetic resonance imaging; PD = pharmacodynamics; TZP = tirzepatide; VAS = visual analog scale.

#### 7.9.1.1. Food Intake (Ad-Libitum Food-Intake Test)

Food intake will be quantified through an ad-libitum food-intake test at lunch in a clinical setting and in free-living conditions on baseline (measured in the lead-in period between Day -4 to Day -1), treatment Week 3 (Visit 6, Day 16) and Week 6 (Visit 11, Day 37), and during the follow-up period specified in the above table.

#### 7.9.1.2. Central Reward and Appetite Circuits using BOLD fMRI

To characterize the treatment effect on central reward and appetite circuits in the brain, MRI scans will be completed during the study. In particular, BOLD fMRI will be completed by subjects at lead-in (measured between Day -4 to Day -1) and treatment Week 3 (Visit 6, Day 16) and Week 6 (Visit 11, Day 37), respectively.

On each of the 3 scheduled visits, 1 MRI scan session will be completed in a fasted state. Subjects will perform a food image task during each scan.

The principal brain reward and appetite areas to be examined will be the

- insula

- medial frontal gyrus
- superior temporal gyrus
- precentral gyrus, and
- cingulate gyrus.

In addition, exploratory brain reward and appetite areas to be examined will be the

- hippocampus
- putamen
- orbitofrontal cortex, and
- ventral striatum.

#### 7.9.1.3. Appetite Visual Analog Scale and Retrospective Appetite Visual Analog Scale

The aim of the appetite VAS is to determine the effects of study treatments on appetite sensations and desire for specific foods. Two types of VASs will be used to assess subjective ratings of appetite and affect: laboratory-based and retrospective. The VASs will be analyzed as continuous variables on the 0 to 100 scale for individual components.

Visual analog scale scores will be documented in an electronic clinical report form (eCRF). During the laboratory-based test meals at baseline (that is, lead-in, measured between Day -4 to Day -1), Week 3, and Week 6, subjects will complete pen-and-paper VAS ratings immediately before and after the test lunch and document scores in the eCRF.

#### 7.9.1.4. Measures of Food Intake-Related and Impulsiveness Questionnaires

The FCI, FCQ-S, Eating Inventory (also known as the TFEQ), Power of Food Scale, and BIS questionnaires are administered at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 postprandially and will be documented in the eCRF.

**Food Craving Inventory**

The 33-item FCI (White et al. 2002) will be used to measure cravings for specific food groups. The measure consists of 5 empirically derived factors. The FCI will be completed at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 in a fasted state shortly after arriving to the study site and will be documented in the eCRF.

**Food Craving Questionnaire-State**

The FCQ-S (Cepeda-Benito et al. 2000; Moreno et al. 2008) is a 15-item measure that assesses the strength of state food cravings at the moment of administration. All items are scored using a 5-point Likert scale in the following manner:

- Strongly disagree - (1).
- Disagree - (2).
- Neutral - (3).
- Agree - (4).
- Strongly agree - (5).

Greater scores for each subscale denote higher levels of craving.

The FCQ-S will be completed at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 in a fasted state shortly after arriving to the study site and will be documented in the eCRF.

**Eating Inventory**

The Eating Inventory (Stunkard and Messick 1985; Stunkard and Messick 1988) or the TFEQ is a 51-item, validated questionnaire that assesses 3 eating-related constructs:

- dietary restraint (21 items)
- disinhibition (16 items), and
- perceived hunger (14 items).

Dietary restraint refers to the intent and ability to restrict food intake, disinhibition measures the tendency to overeat, and perceived hunger measures susceptibility to feelings of hunger. Different items have different scales. The TFEQ will be completed at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 and will be documented in the eCRF.

**Power of Food Scale**

The 21-item Power of Food Scale (Cappelleri et al. 2009; Lowe et al. 2009) measures appetite for palatable foods. Among 21 items, 15 items are used to calculate the subdomain scores at the following 3 levels of food proximity (3 domain scores):

- food available (6 items)
- food present (4 items), and
- food tasted (5 items).

The Power of Food Scale will be completed at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 and will be documented in the eCRF.

**Barratt Impulsiveness Scale**

Impulsivity will be measured with the BIS. The BIS (Patton et al. 1995; Reid et al. 2014) is a 30-item, self-reported measure describing impulsive or nonimpulsive behaviors and preferences.

The BIS will be completed at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 in a fasted state shortly after arriving to the study site and will be documented in the eCRF.

#### 7.9.1.5. Hormones and Metabolites Related to Appetite Regulation

To explore potential effects of study treatment on hormones related to appetite and the regulation of nutrient metabolism, plasma or serum samples will be collected and assayed. The following hormones measured at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 at a central laboratory will be exploratory PD parameters of interest and will be only tested once from either plasma or serum:

- amylin (postprandial)
- ghrelin (fasting)
- GIP (postprandial)
- GLP-1 (postprandial)
- leptin (postprandial)
- pancreatic polypeptide (postprandial)
- PYY (postprandial)
- insulin (fasting and postprandial), and
- glucagon (fasting).

In addition to safety monitoring, the following metabolites will also be measured as exploratory endpoints at baseline (Day -4; local laboratory), Week 3 (central laboratory), and Week 6 (central laboratory):

- plasma glucose (fasting), and
- triglycerides (fasting).

Glucose and triglycerides do not have central laboratory measures at baseline, so local laboratory measures will be used for baseline.

## *7.9.2. Pharmacodynamic Statistical Inference*

Pharmacodynamic parameters at baseline (measured in the lead-in period between Day -4 to Day -1), Week 3, and Week 6 (if applicable), as well as change from baseline values at Week 3 and Week 6 and change from Week 3 at Week 6, will be summarized by treatment. Plots of mean PD parameters and mean changes from baseline over time may be provided by treatment and by time (that is, Week) if deemed necessary. The change from baseline to Week 3 and Week 6 (if applicable), calculated for all the continuous PD parameters, will be analyzed.

### 7.9.2.1. Primary Pharmacodynamic Analysis

#### *7.9.2.1.1. Energy Intake via Ad-libitum Food Intake Test*

The primary endpoint in this study is the change from baseline to Week 3 in energy intake (kcal) as assessed by an ad-libitum food-intake test at lunch in a clinical setting for the comparison of 5-mg tirzepatide versus placebo.

For clinical settings, the amount of grams of fat (saturated and unsaturated), carbohydrates, and protein consumed will be documented in the eCRF. These calories will be represented in kcal format using the following conversion: 1 gram of carbohydrate or protein = 4 kcal, 1 gram of saturated fat or unsaturated fat = 9 kcal. The energy intake will be calculated as the total calories (kcal) and recorded in an eCRF as well. The percentage of fat, carbohydrate, and protein in total energy may be derived:

- food intake
  - g (that is, the amount of grams of food consumed)
- fat (saturated and unsaturated)
  - g and kcal

  - percent of energy (that is, fat [kcal]/energy intake [kcal])
- carbohydrate
  - g and kcal
  - percent of energy (that is, carbohydrate [kcal]/energy intake [kcal]), and
- protein
  - g and kcal
  - percent of energy (that is, protein [kcal]/energy intake [kcal]).

The primary endpoint will be analyzed using an analysis of covariance to compare the effect of tirzepatide versus placebo at Week 3 on lunch energy intake (kcal) with treatment as fixed effect, baseline BMI stratum, and baseline energy (kcal) intake as covariates in the model using the PD analysis set.

Inferential statistics will include LSMs (Standard error) of energy (kcal) intake by treatment and the treatment difference of tirzepatide versus placebo in LS Mean difference with 95% CI, and the p value for treatment comparison will be displayed.

#### 7.9.2.2. Secondary Pharmacodynamic Analyses

Secondary PD measures (that is, BOLD fMRI activation; VAS; and the FCI, FCQ-S, TFEQ, and Power of Food Scale questionnaires [see Section 7.10.1 for details]) are scheduled to be measured twice postbaseline at Week 3 and Week 6.

The PD measurements will be analyzed using a MMRM method to compare the effect of tirzepatide versus placebo at Week 3. The response variable is change from baseline in the PD measurement (that is, postbaseline - baseline), where postbaseline measurement is collected at Week 3 and Week 6. The model will include treatment, baseline BMI stratum, week, and treatment-by-week interaction as fixed effects and baseline PD measurement as a covariate; scanner ID will be included as a covariate for modelling the function MRI parameters as well. An unstructured covariance structure will be used to model the within-subject variabilities if deemed appropriate. The restricted-maximum-likelihood approach will be used to obtain model parameter estimates.

For both secondary and exploratory PD parameters, a single model for each parameter will be constructed. The inferential results will be displayed into section for secondary endpoints and section for exploratory endpoints, separately.

For secondary endpoints, inferential statistics will include LSMs (Standard error) of each treatment, the treatment difference with 95% CI at Week 3, along with the p value for comparison.

##### *7.9.2.2.1. Central Reward and Appetite Circuits in the Fasting State using BOLD fMRI*

Secondary PD measures to assess the effect of 5-mg tirzepatide versus placebo at Week 3 on parameters of central reward and appetite circuits during fasting and postprandial states using BOLD fMRI are

- change from baseline in mean difference to Week 3 in BOLD fMRI activation to images of highly palatable food (versus nonfood objects) during fasting states in 5 brain reward and appetite areas (insula, medial frontal gyrus, superior temporal gyrus, precentral gyrus, and cingulate gyrus).

The 5 principal brain reward and appetite ROI to be examined and recorded will be

- insula
- medial frontal gyrus
- superior temporal gyrus
- precentral gyrus, and
- cingulate gyrus.

Four additional exploratory brain reward and appetite ROI to be examined and recorded will be

- hippocampus
- putamen
- orbitofrontal cortex, and
- ventral striatum.

Mean differences in BOLD signal between views of highly palatable food (high-fat, high-sugar and high-fat, high-carbohydrate) versus nonfood objects (for example, office supplies and furniture) will be calculated within 5 principal food-reward and appetite-related ROI and 4 exploratory food-reward and appetite-related ROI.

These validated contrast values (that is, mean difference in BOLD signal) (continuous) will be provided by image data processing vendor (Clario) to statisticians/programmers for analyses. Change from baseline in BOLD contrast values at Week3 and at Week6 will be summarized by treatment.

Additional analyses using other functional contrasts (for example, highly palatable foods versus not highly palatable foods [low-fat, low-carbohydrate, high-protein] and not highly palatable foods versus nonfood objects) may be conducted at the 5 principal food-reward and appetite-related ROI and 4 exploratory food-reward and appetite-related ROI if deemed necessary.

#### *7.9.2.2.2. Behavioral Appetite Assessments*

Secondary PD measures to assess the effect of 5-mg tirzepatide versus placebo at Week 3 on parameters of behavioral appetite assessments are

- change from baseline to Week 3
  - in fasting and postprandial appetite VAS (laboratory-based), and
  - fasting FCI, FCQ-S, TFEQ, and Power of Food Scale questionnaires.

The parameters will be summarized for each at lead-in, Week 3, and Week 6, and the change from baseline at Week 3 and Week 6 will be also summarized. The comparison of secondary objective parameters will be analyzed using an MMRM model respectively for each parameter.

- Response variable - Change from baseline in a secondary PD parameter.

- Independent variables - Treatment, time, and treatment-by-time interaction as fixed effects; baseline value in secondary parameter as a covariate; and patient as a random effect.

Inferential statistics include LSMs and standard error of each of the secondary objective parameters for tirzepatide and placebo, and the estimated treatment difference (tirzepatide versus placebo) at Week 3, corresponding 2-sided 95% CI, and associated p-values.

#### 7.9.2.2.2.1. Fasting and Postprandial Appetite VAS (Laboratory-Based)

Secondary PD measures to compare the treatment effect of 5-mg tirzepatide versus placebo at Week 3 are

- change from baseline (lead-in) in overall score and individual score of fasting and postprandial appetite VAS in laboratory-based setting, and
- change from baseline (lead-in) in overall score and individual score of retrospective Appetite VAS.

The aim of the VAS is to determine the effects of study treatments on appetite sensations and desire for specific foods. Two types of VASs will be used to assess subjective ratings of appetite and effect: laboratory-based appetite VAS and retrospective appetite VAS. The VAS scales will be analyzed as continuous variables on the 0-to-100 scale for individual components. The VAS scores will be documented in eCRF.

The laboratory-based appetite VAS measures subjects at lead-in, Week 3, Week 6, and immediately before and immediately after the ad-libitum food intake test. The 8 individual questions measure hunger, satiety, fullness, prospective food consumption, desire for sweet food, desire for salty food, desire for savory food, and desire for fatty food. For the first 4 ratings, 0 means “Not at all” and 100 means “Extremely”. For the last 4 ratings, 0 means “Yes, very much” and 100 means “No, not at all”.

The retrospective appetite VAS (Womble et al. 2003) will be used to measure the average ratings of appetite that subjects experienced over the past week. In addition to the 4 appetite metrics (satiety, fullness, prospective food consumption, and hunger) noted above (with instructions modified), retrospective VASs will also be used to quantify feelings of nausea, malaise, and gastrointestinal distress over the previous week. These ratings will occur at baseline, Week 3, and Week 6.

Overall appetite score is calculated as the average of the first 4 individual scores (van Can et al. 2014; Flint et al. 2000):

$$\textit{overall appetite score} = (\textit{satiety} + \textit{fullness} + [100 - \textit{prospective food consumption}] + [100 - \textit{hunger}]) / 4$$

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

Individual scores (satiety, fullness, hunger, prospective food consumption, desire for sweet food, desire for salty food, desire for savory food, and desire for fatty food) and overall appetite scores

for the Appetite VAS and individual scores (satiety, fullness, hunger, prospective food consumption, nausea, malaise, and gastrointestinal distress) and overall appetite scores of 4 appetite metrics for the retrospective VAS will be summarized at baseline, Week 3, and Week 6, and the change from baseline at Week 3 and Week 6 will also be summarized.

The secondary objective of comparing the treatment effect of 5-mg tirzepatide versus placebo at Week 3 in the change from baseline (lead-in) in the overall score of fasting and postprandial laboratory appetite VAS and retrospective appetite VAS will be analyzed with MMRM model in a way mentioned in the beginning of this section. The change from baseline (lead-in) in overall score of four 4 appetite metrics for the retrospective VAS will be also analyzed with an MMRM model similarly.

#### 7.9.2.2.2.2. Fasting FCI

A secondary objective to compare the effect of 5-mg tirzepatide versus placebo at Week 3 is change from baseline in fasting FCI

The 33-item FCI (White et al. 2002) is used to measure cravings for specific 5 groups (high fats, sweets, carbohydrates/starches, fast food fats, and fruits and vegetables). 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 questionnaire was completed. All items are scored in the following 5-point Likert scale

- never - 1.
- rarely - 2.
- sometimes - 3.
- often - 4.
- always - 5.
    - Greater scores denote higher levels of craving.

Secondary objectives to compare the effect of 5-mg tirzepatide versus placebo at Week 3 for the change from baseline in fasting FCI including 5 subscores (average of nonmissing measurements of individual items in that category) and overall scores (average of nonmissing measurements of all individual items) are

- **Subscores**
    - High fats (8 items) - Fried chicken, gravy, sausage, hot dog, fried fish, corn bread, bacon, and steak.
    - Sweets (8 items) - Cake, cinnamon rolls, ice cream, cookies, chocolate, donuts, candy, and brownies.
    - Carbohydrates/starches (8 items) - Sandwich bread, rice, biscuits, pasta, pancakes or waffles, rolls, cereal, and baked potato.
    - Fast food fats (4 items) - Pizza, french fries, hamburger, and chips.
    - Fruits and vegetables (5 items) - Cooked vegetables, fruit juices, raw vegetables, canned fruit, and raw fruit.
- **Overall score** - All 33 items.

The subscale score for each factor and overall score will be summarized at lead-in, Week 3, and Week 6. Change from baseline (lead-in) will be also summarized for the Week 3 and Week 6 visits. Change from Week 3 at Week 6 will be summarized as well. Mean plot and mean change from baseline plot for each subscore and overall score by visit will be supplied.

The secondary objective of comparing the treatment effect of tirzepatide versus placebo at Week 3 in FCI overall score will be analyzed using an MMRM model in a way mentioned in the beginning of this section.

#### 7.9.2.2.2.3. FCQ-S

A secondary objective to compare the treatment effect 5-mg tirzepatide versus placebo at Week 3 is change from baseline in FCQ-S, including 5 subscores and the overall score in the FCQ-S questionnaire.

The FCQ-S (Cepeda-Benito et al. 2000; Moreno et al. 2008) is a 15-item measure that assesses the strength of state food cravings at the moment of administration. The questionnaire contains 5 subscale scores, including desire, anticipation positive, anticipation negative, lack of control, and hunger.

**Table GPHH.7.3. FCQ-S Subscore Category and Questions**

| Subscore category | Question Number and Description |
|---|---|
| Desire (3 items) | 1. I have an intense desire to eat [one or more specific foods]. |
| | 2. I'm craving. |
| | 3. I have an urge for [one or more specific foods]. |
| Anticipation positive (3 items) | 4. Eating [one or more specific foods] would make things seem just perfect. |
| | 5. If I were to eat what I am craving, I am sure my mood would improve. |
| | 6. Eating [one or more specific foods] would feel wonderful. |
| Anticipation negative (3 items) | 7. If I ate something, I wouldn't feel so sluggish and lethargic. |
| | 8. Satisfying my craving would make me feel less grouchy and irritable |
| | 9. I would feel more alert if I could satisfy my craving. |
| Lack of control (3 items) | 10. If I had [one or more specific foods], I could not stop eating it. |
| | 11. My desire to eat [one or more specific foods] seems overpowering. |
| | 12. I know I'm going to keep on thinking about [one or more specific foods] until I actually have it. |
| Hunger (3 items) | 13. I am hungry. |
| | 14. If I ate right now, my stomach wouldn't feel as empty. |
| | 15. I feel weak because of not eating. |

All items are scored using a 5-point Likert scale in the following manner:

- Strongly disagree - 1.
- Disagree - 2.
- Neutral - 3.
- Agree - 4.
- Strongly agree - 5.
    - Greater scores for each subscale denote higher levels of craving.
- **Subscores -** average score from a nonmissing measurement from an individual question.
    - Desire (average score from questions 1-3)
    - Anticipation positive (average of questions 4-6)
    - Anticipation negative (average of questions 7-9)
    - Lack of control (average of questions 10-12), and
    - Hunger (average of questions 13-15).
- **Overall score - Average of the 5 subscores above.**

Each subscale score and overall score will be summarized for the lead-in, Week 3, and Week 6 visits; change from lead-in will be also summarized for the Week 3 and Week 6 visits. Change from Week 3 at Week 6 will be summarized as well. Mean plot and individual plots for each subscale score by visit will be provided.

The secondary objective of comparing the treatment effect of tirzepatide versus placebo at Week 3 in FCQ-S overall score will be analyzed using an MMRM model in a way mentioned in the beginning of this section.

#### 7.9.2.2.2.4. Eating Inventory through TFEQ

A secondary objective to compare the treatment effect 5-mg tirzepatide versus placebo at Week 3 is change from baseline in the TFEQ, including 3 factors:

- dietary restraint
- disinhibition, and
- perceived hunger.

Of the 51 items of the TFEQ, 36 items are on a binary-point (true or false) response scale, 1 item is on a 6-point response scale, and 14 items are on a 4-point response scale, respectively. Scores from sum of individual items for dietary restraint, disinhibition, and perceived hunger range from 0 to 21, 0 to 16, and 0 to 14, respectively, and a greater number indicates greater levels of each respective eating component:

- Binary point questions (36)-1-36
- 4-point responses questions (14)-37-49
- 6-point response questions (1)-51

**Table GPHH.7.4. Scoring Algorithm for the TFEQ**

| Factor | Binary-Point (True/False) Question No. | 6-Point or 4-Point Response Scale Question No. |
|---|---|---|
| 1 - Dietary restraint (21 items) | 4, 6, 10, 14, 18, 21, 23, 28, 30, 32, 33, and 35<br><br>**To score 1 point:**<br>• **Answer true -** 4, 6, 14, 18, 23, 28, 32, 33, 35<br>• **Answer false -** 10, 21, 30 | 37, 38, 40, 42, 43, 44, 46, 48, and 51<br><br>**To score 1 point for 37, 38, 40, 42, 43, 44, 46, or 48**<br>• Selected the third or fourth choice<br>**To score 1 point for 51**<br>• Selected the fourth, fifth, or sixth choice |
| 2 - Disinhibition (16 items) | 1, 2, 7, 9, 11, 13, 15, 16, 20, 25, 27, 31, and 36<br><br>**To score 1 point**<br>• **Answer true** - 1, 2, 7, 9, 11, 13, 15, 20, 27, 36<br>• **Answer false** - 16, 25, 31 | 45, 49, and 50<br><br>**To score 1 point**<br>• Selected the third or fourth choice |
| 3 - Perceived hunger (14 items) | 3, 5, 8, 12, 17, 19, 22, 24, 26, 29, 34<br><br>**To score 1 point**<br>• Answer true - 3, 5, 8, 12, 17, 19, 22, 24, 26, 29, 34 | 39, 41, and 47<br><br>**To score 1 point for 39 or 41**<br>• Selected the third or fourth choice<br>**To score 1 point for 47**<br>• Selected the first or second choice |

Abbreviations: No. = number; TFEQ = Three-Factor Eating Questionnaire.

1) Unless specified above, for any other answer/selection, the patient will score 0 for that question item.

**2) For question 51**, the scale is 1, 2, 3, 4, 5, and 6 sequentially. If subject selected the fourth choice (scale = 4), the fifth choice(scale = 5), or the sixth choice (scale = 6), the subject will score 1 point from this question.

3) The complete questionnaire will be supplemented in appendix.

The TFEQ will be completed at lead-in, Week 3, and Week 6 in a fasted state shortly after arriving to the study site and will be documented in the eCRF.

Summaries of each factor score will be provided for lead-in, Week 3, and Week 6. Change from lead-in at Week 3 and Week 6, and change from Week 3 at Week 6, in each factor score will be summarized by treatment group. Individual and mean plots will be provided.

The secondary objective of comparing the treatment effect of tirzepatide versus placebo at Week 3 in each TFEQ factor score will be analyzed using an MMRM model in a way mentioned in the beginning of this section.

#### 7.9.2.2.2.5. Power of Food Scale Questionnaires

Secondary objectives of comparing the treatment effect of tirzepatide versus placebo at Week 3 are

- change from baseline in each domain score of the Power of Food Scale Questionnaire
  - Food Available Domain
  - Food Present Domain, and

  - Food Tasted Domain, and
- change from baseline in aggregate score of PFS (mean of the 3 domain scores).

The 21-item Power of Food Scale (Cappelleri et al. 2009; Lowe et al. 2009) measures appetite for palatable foods in the environment at the following 3 levels of food proximity (3 domain scores): food available, food present, and food tasted. The original questionnaire was designed with 21 items, and then 6 items were moved after confirmatory factor analysis. In Study I8F-MC-GPHH, all 21 items will be collected from an eCRF, but only 15 of those will be used to calculate domain scores.

**Table GPHH.7.5. Power of Food Scale Subdomains and Questions**

| Domain | Question Number (based on 21-item) | Question |
|---|---|---|
| Food Available Domain (6 items) | 1 | I find myself thinking about food even when I'm not physically hungry. |
| | 3 | I get more pleasure from eating than I do from almost anything else. |
| | 8 | It's scary to think of the power that food has over me. |
| | 16 | Sometimes, when I am doing everyday activities, I get an urge to eat "out of the blue" (for no apparent reason). |
| | 17 | I think I enjoy eating a lot more than most other people. |
| | 19 | It seems like I have food on my mind a lot. |
| Food Present Domain (4 items) | 5 | If I see or smell a food I like, I get a powerful urge to have some. |
| | 6 | When I'm around fattening food I love, it's hard to stop myself from at least tasting it. |
| | 10 | When I know a delicious food is available, I can't help myself from thinking about having some. |
| | 11 | I love the taste of certain foods so much that I can't avoid eating them even if they are bad for me. |
| Food Tasted Domain (5 items) | 14 | Just before I taste a favorite food, I feel intense anticipation. |
| | 15 | When I eat delicious food I focus a lot on how good it tastes. |
| | 18 | Hearing someone describe a great meal makes me really want to have something to eat. |
| | 20 | It is very important to me that the foods I eat are as delicious as possible. |
| | 21 | Before I eat a favorite food my mouth tends to flood with saliva. |

The aggregate score is also calculated as the mean of the 3 domain scores. All items are scored on a 5-point Likert scale in the following manner:

- Don't agree at all - 1.
- Agree a little - 2.
- Agree somewhat - 3.
- Agree - 4.
- Strongly agree - 5.
  - A higher item score indicates a greater responsiveness to the food environment.

The aggregate score and each of the 3 domain scores will be summarized for lead-in, Week 3, and Week 6, and the change from lead-in will be summarized for the Week 3 and Week 6 visits.

The secondary objective of comparing the treatment effect of tirzepatide versus placebo at Week 3 in each domain score and aggregate score of the Power of Food Scale Questionnaire will be analyzed using an MMRM model in a way mentioned in the beginning of this section.

#### 7.9.2.2.2.6. Barratt Impulsiveness Scale

The secondary objective of comparing the treatment effect of Tirzepatide versus Placebo on impulsivity at Week 3 is change from baseline in each of the 6 first-order subscores and each of the 3 second-order subscores in the BIS.

Impulsivity will be measured with the BIS. The BIS (Patton et al. 1995; Reid et al. 2014) is a 30-item, self-reported measure describing impulsive or nonimpulsive behaviors and preferences. The measure assesses the personality/behavioral construct of impulsiveness and has demonstrated good psychometric properties in multiple populations. The items are rated on a 4-point scale with responses of

- rarely/never
- occasionally
- often, and
- almost always/always.

The maximum score of 4 indicates the most impulsive response.

The scale consists of 6 first-order factors, and the subscores are calculated as follows:

- Attention (Factor 1) - Average of 5, 9, 11, 20, and 28.
- Motor impulsiveness (Factor 2) - Average of 2, 3, 4, 17, 19, 22, and 25.
- Self-control (Factor 3) - Average of 1, 7, 8, 12, 13, and 14.
- Cognitive complexity (Factor 4) - Average of 10, 15, 18, 27, and 29.
- Perseverance (Factor 5) - Average of 16, 21, 23, and 30.
- Cognitive instability (Factor 6) - Average of 6, 24, and 26.

The scale also consists of 3 second-order factors:

- Attentional impulsiveness (Factor I) - Average of 5, 6, 9, 11, 20, 24, 26, and 28.
- Motor impulsiveness (Factor II) - Average of 2, 3, 4, 16, 17, 19, 21 22, 23, 25, and 30.
- Non-planning Impulsiveness (Factor III) - Average of 1, 7, 8, 10, 12, 13, 14, 15, 18, 27, and 29.

The subscore for a specific factor, first or second order, is calculated as the average of the corresponding items. A total score is calculated as the average of all items.

**Table GPHH.7.6. Questions Categorized into First-Order Categories**

| First-Order Category | Item Number | Item Description |
|---|---|---|
| 1 - Attention (5 items) | 5 | I don't "pay attention". |
| | 9 | I concentrate easily. |
| | 11 | I "squirm" at plays or lectures. |
| | 20 | I am a steady thinker. |
| | 28 | I am restless at the theater or lectures. |
| 2 - Motor impulsiveness (7 items) | 2 | I do things without thinking |
| | 3 | I make-up my mind quickly. |
| | 4 | I am happy-go-lucky. |
| | 17 | I act "on impulse". |
| | 19 | I act on the spur of the moment. |
| | 22 | I buy things on impulse. |
| | 25 | I spend or charge more than I earn. |
| 3 - Self-control (6 items) | 1 | I plan tasks carefully |
| | 7 | I plan trips well ahead of time. |
| | 8 | I am self controlled. |
| | 12 | I am a careful thinker. |
| | 13 | I plan for job security. |
| | 14 | I say things without thinking. |
| 4 - Cognitive complexity (5 items) | 10 | I save regularly. |
| | 15 | I like to think about complex problems. |
| | 18 | I get easily bored when solving thought problems. |
| | 27 | I am more interested in the present than the future. |
| | 29 | I like puzzles. |
| 5 - Perseverance (4 items) | 16 | I change jobs. |
| | 21 | I change residences. |
| | 23 | I can only think about one thing at a time. |
| | 30 | I am future oriented. |
| 6 - Cognitive instability (3 items) | 6 | I have "racing" thoughts. |
| | 24 | I change hobbies. |
| | 26 | I often have extraneous thoughts when thinking. |

**Table GPHH.7.7. Questions Categorized into Second-Order Categories**

| Second-Order Category | Item Number |
|---|---|
| I - Attentional impulsiveness (8 items) | 5, 6, 9, 11, 20, 24, 26, and 28 |
| II - Motor impulsiveness (11 items) | 2, 3, 4, 16, 17, 19, 21 22, 23, 25, and 30 |
| III - Non-planning impulsiveness (11 items) | 1, 7, 8, 10, 12, 13, 14, 15, 18, 27, and 29 |

The BIS will be completed at lead-in, Week 3, and Week 6 in a fasted state shortly after arriving to the study site and will be documented in the eCRF.

Each of the 6 first-order subscores, each of the 3 second-order subscores, and the total score from the BIS questionnaire will be summarized at lead-in, Week 3, and Week 6, and the change from lead-in will be summarized for the Week 3 and Week 6 visits.

The secondary objective of comparing the treatment effect of tirzepatide versus placebo at Week 3 in each of the 6 first-order subscores, each of the 3 second-order subscores, and the total score from the BIS questionnaire will be analyzed using an MMRM model in a way mentioned in the beginning of this section.

#### 7.9.2.3. Exploratory Pharmacodynamic Analyses

The PD measurements will be analyzed using the same MMRM method used for secondary PD parameters to compare the effect of tirzepatide versus placebo at Week 3 and/or Week 6.

Inferential statistics will include LSMs of each treatment, the treatment difference, the standard error, and 95% CI at Week 3 and/or Week 6 for below specified exploratory parameters.

**Central reward and appetite circuits in the fasting state using BOLD fMRI**

An exploratory PD measure to compare the effect of 10-mg tirzepatide versus placebo at Week 6 on parameters of central reward and appetite circuits during fasting state using BOLD fMRI is

- change from baseline to Week 6 in BOLD fMRI activation to images of highly palatable food (high fat-high sugar and high fat-high carbohydrate) (versus nonfood objects) during the fasting states in **5** brain reward and appetite areas (insula, medial frontal gyrus, superior temporal gyrus, precentral gyrus, and cingulate gyrus).

An exploratory PD measure to compare the effect of 5- and 10-mg tirzepatide versus placebo at Week 3 and Week 6 on parameters of central reward and appetite circuits during fasting state using BOLD fMRI is

- change from baseline to Week 3 and Week 6 in BOLD fMRI activation to images of highly palatable food (high fat-high sugar and high fat-high carbohydrate) (versus nonfood objects) during the fasting states in **4** brain reward and appetite areas (hippocampus, putamen, orbitofrontal cortex, and ventral striatum).

An exploratory PD measure to compare the effect of 5-mg and 10-mg tirzepatide versus liraglutide and liraglutide versus placebo at Week 3 and Week 6 on parameters of central reward and appetite circuits during fasting state using BOLD fMRI is

- change from baseline to Week 3 and Week 6 in BOLD fMRI activation to images of highly palatable food (high fat-high sugar and high fat-high carbohydrate) (versus nonfood objects) during the fasting states in **9** brain reward and appetite areas (insula, medial frontal gyrus, superior temporal gyrus, precentral gyrus, cingulate gyrus, hippocampus, putamen, orbitofrontal cortex, and ventral striatum).

Guided by results, additional analyses on parameters of central reward and appetite circuits using other functional contrasts (for example, highly palatable foods versus not highly palatable foods, not highly palatable foods versus nonfood objects) may be conducted at the 5 principal food-reward and appetite-related ROI and 4 exploratory food-reward and appetite-related ROI as deemed appropriate.

**Behavioral appetite assessments**

Exploratory PD measures to compare the effect of 10-mg tirzepatide versus placebo at Week 6 on parameters of behavioral appetite assessments are

- change from baseline to Week 6
  - in fasting and postprandial appetite VAS (laboratory-based), and
  - in the fasting FCI, FCQ-S, TFEQ, and Power of Food Scale questionnaires.

Exploratory PD measures to compare the effect of 5-mg and 10-mg tirzepatide versus liraglutide and liraglutide versus placebo at Week 3 and Week 6 on parameters of behavioral appetite assessments are

- change from baseline to Week 3 and Week 6
  - in fasting and postprandial appetite VAS (laboratory-based), and
  - in the fasting FCI, FCQ-S, TFEQ, and Power of Food Scale questionnaires.

**Energy intake (ad-libitum food-intake test)**

An exploratory PD measure to compare the effect of 10-mg tirzepatide versus placebo at Week 6 on energy intake in a clinic-based setting is

- change from baseline in energy intake (kcal) at lunch to Week 6 as assessed in a clinic-based setting.

An exploratory PD measure to compare the effect of 5-mg and 10-mg tirzepatide versus liraglutide and liraglutide versus placebo at Week 3 and Week 6 on energy intake in a clinic-based setting is

- change from baseline in energy intake (kcal) at lunch to Week 3 and Week 6 as assessed in a clinic-based setting.

**Impulsivity**

An exploratory PD measure to compare the effect of 5-mg and 10-mg tirzepatide versus placebo at Week 3 and Week 6 on parameters of Impulsivity through BIS questionnaire is

- change from baseline in impulsivity in BIS questionnaire.

**Hormones and metabolites related to appetite regulation**

To explore potential effects of study treatment on hormones related to appetite and the regulation of nutrient metabolism, plasma or serum samples will be collected and assayed. The following hormones measured at baseline, Week 3, and Week 6 at a central laboratory will be exploratory PD parameters of interest and will be only tested once from either plasma or serum:

- amylin (postprandial)
- ghrelin (fasting)
- GIP (postprandial)
- GLP-1 (postprandial)
- leptin (postprandial)
- pancreatic polypeptide (postprandial)
- PYY (postprandial)
- insulin (fasting and postprandial), and
- glucagon (fasting).

In addition to safety monitoring, the following metabolites will also be measured as exploratory endpoints at lead-in (Day -4; local laboratory), Week 3 (central laboratory), and Week 6 (central laboratory): plasma glucose (fasting) and triglycerides (fasting). Note that glucose and triglycerides do not have central laboratory measures at baseline, so local laboratory measures will be used for baseline.

Exploratory PD measures to compare the effect of 5- and 10-mg tirzepatide versus placebo at Week 3 and Week 6 on hormones and metabolites related to appetite regulation are

- change from baseline to Week 3 and Week 6 in
    - ghrelin (fasting)
    - glucagon (fasting)
    - insulin (fasting and postprandial)
    - amylin (postprandial)
    - GIP (postprandial)
    - GLP-1 (postprandial)
    - leptin (postprandial)
    - pancreatic polypeptide (postprandial), and
    - PYY (postprandial)
- change from baseline to Week 3 and Week 6 in plasma glucose (fasting) and triglycerides (fasting).

Exploratory PD measures to compare the effect of 5-mg and 10-mg tirzepatide versus liraglutide, and liraglutide versus placebo, at Week 3 and Week 6 on hormones and metabolites related to appetite regulation are

- change from baseline to Week 3 and Week 6 in
    - ghrelin (fasting)
    - glucagon (fasting)
    - insulin (fasting and postprandial)
    - amylin (postprandial)
    - GIP (postprandial)
    - GLP-1 (postprandial)
    - leptin (postprandial)
    - pancreatic polypeptide (postprandial), and

    - PYY (postprandial)
- change from baseline to Week 3 and Week 6 in plasma glucose (fasting) and triglycerides (fasting).

Hormone and metabolites analyses are considered as exploratory analyses only. The estimand is pairwise treatment difference in terms of change from baseline to treatment Week 3 and 6. In particular, interest lies in comparing tirzepatide 5 mg with liraglutide, tirzepatide 10 mg with liraglutide, and liraglutide with placebo. All 11 variables described above will be analyzed separately using an MMRM model.

**Summary for PD analysis**

For PD analyses, when the distribution of response variables significantly deviates from normality, alternative statistical methods will be utilized.

Guided by results, for BOLD fMRI , additional analyses on parameters of central reward and appetite circuits using other functional contrasts (for example, highly palatable foods versus not highly palatable foods, not highly palatable foods versus nonfood objects) may be conducted at the 5 principal food-reward and appetite-related ROI and 4 exploratory food-reward and appetite-related ROI as deemed appropriate.

Inferential statistics will include LSMs of each treatment, the treatment difference, standard error, and 95% CI.

## 7.10. Safety Analyses

Safety measures include, but are not limited to, AEs, TEAEs, SAEs, AESI, vital signs, and safety laboratory measures.

Safety analyses will be performed on the safety population unless specified otherwise. Safety listings will display values/events during study period. Listings of AEs, deaths, and SAEs may include (but will not be limited to)

- subject identification (ID) number
- age
- sex
- race
- treatment
- dose
- MedDRA SOC and PT
- 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 deemed appropriate.

Additional safety listings will be provided for safety parameters other than AEs if specified in the sections below.

For safety measurements, summary statistics will be provided by treatment. A summary for AEs with a frequency of 10 or more patients with such events will be provided as well.

### *7.10.1. Adverse Events*

Adverse events for the safety population will be listed, which will include the MedDRA PT.

### *7.10.2. Treatment-Emergent Adverse Events*

A TEAE is defined as an AE that first occurs post first dose or that is present prior to the 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 AEs will be summarized by treatment, severity, and relationship with study drug.

### *7.10.3. Serious Adverse Events*

A listing of patients with SAEs, including deaths, will be produced.

### *7.10.4. Adverse Events Leading to Discontinuation*

A listing of patients with AEs leading to discontinuation from study by treatment will be presented. A summary will be provided if there are 10 or more subjects with such events.

### *7.10.5. Special Safety Topics*

A listing of patients with all AESI defined in Section 7.10.5 will be provided.

#### 7.10.5.1. Hypoglycemia

A listing of documented, clinically significant hypoglycemia (plasma glucose less than 54 mg/dL) and severe hypoglycemia events will be provided. A listing of patients with hypoglycemic events will be provided. The category of hypoglycemic events will be presented. Hypoglycemic events will be summarized by treatment for each category. The incidence of hypoglycemia will be reported.

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

**Table GPHH.7.8. Definition 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** is 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** is defined as any hypoglycemia events that occurs between bedtime and waking.

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

### 7.10.5.2. Pancreatitis

A listing of patients with (acute) pancreatic events will be provided.

Treatment-emergent, investigator assessed pancreatitis will be considered as an AESI.

### 7.10.5.3. Thyroid Malignancies and C-Cell Hyperplasia

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

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

### 7.10.5.4. Hypersensitivity Events

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

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

#### 7.10.5.5. Injection-Site Reactions

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

A listing of patients with reported injection-site reactions (edema, erythema, induration, itching, and pain) will be provided. A frequency table by treatment of the number of subjects reporting 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 (for example, abscess, cellulitis, erythema, hematomas/hemorrhage, exfoliation/necrosis, pain, subcutaneous nodules, swelling, indurating, and inflammation) will be considered as AESI.

#### 7.10.5.6. 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 AESI.

**Hepatic monitoring**

The subjects' liver disease history and associated data will be listed. Any concomitant medications that have potential for hepatotoxicity, including acetaminophen, will be listed. Results from any hepatic monitoring procedures, such as a magnetic resonance elastography 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 summarized by treatment and listed. Alcohol and recreational drug-use data will also 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.

#### 7.10.5.7. Severe/Serious Gastrointestinal Adverse Events

A listing of patients 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 AESI.

#### 7.10.5.8. Acute Renal Events

A listing of patients with acute renal events (search criteria in Appendix 1) will be provided. Serious/severe acute renal events will be considered as AESI.

### *7.10.6. Safety Laboratory Parameters*

All laboratory data will be reported in International System of Units and conventional units. Clinical chemistry, hematology and endocrine (calcitonin) data and their changes from baseline will be summarized at each planned visit by treatment using descriptive statistics.

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 QL (0.5 × QL may be used for the calculation of summary statistics); (2) above the QL (1.1 × QL may be used for the calculation of summary statistics, if deemed appropriate).

### 7.10.7. Vital Signs

Descriptive summaries of vital signs by treatment and by visit will be provided for the lead-in, baseline (Day 1 predose), postbaseline, 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 GPHH.7.9.

**Table GPHH.7.9. 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 |

Abbreviation: BP = blood pressure.

### 7.10.8. Body Weight and Waist Circumference

An average of 2 recordings will be derived in weight and waist circumference for each applicable visit. Absolute value and change from baseline (Day 1 predose) will be summarized by treatment for each visit. For change from baseline, a MMRM based model (similar to the MMRM model used for PD parameters) will be used to model the treatment effect in Body Weight and Waist Circumference at Week 3 and Week 6.

### 7.10.9. Electrocardiogram

Electrocardiograms will be performed for safety monitoring purposes only and will not be presented. Any clinically significant findings from ECGs will be recorded as AEs.

## 7.11. Immunogenicity

The baseline immunogenicity sample is collected on Day 1 predose.

The frequency and percentage of subjects with preexisting ADAs and that are treatment-emergent ADA positive (TE-ADA+) to tirzepatide will be tabulated.

Treatment-emergent ADAs are defined as those with a titer 2-fold (1 dilution) greater than the minimum required dilution if no ADAs were detected at baseline (Day 1 predose) (treatment-induced ADA) or those with a 4-fold (2 dilutions) increase in titer compared to baseline if ADAs were detected at baseline (treatment-boosted ADAs). For the TE-ADA+ subjects, the distribution of maximum titers may be described.

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

The relationship between the presence of antibodies and PD response, including safety and efficacy to tirzepatide, may be assessed. In particular, listings will be provided to show the relationship between ADA response and hypersensitivity and injection site reaction.

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

## 7.12. 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 SAP.

## 7.13. COVID-19 Impact Assessment

A listing of patients with AEs related to COVID-19, including death due to COVID-19, serious COVID-19 AEs, and COVID-19 AEs may be provided. Protocol deviations that related to COVID-19 may be described in the clinical study report.

## 7.14. Exploratory Analysis for Motion Sensor Modeling

To evaluate and refine an eating-activity-pattern recognition model, passive detection of eating-activity data will be collected passively via a wearable wrist device during the last week of the follow-up period of the study.

The analyses of the wearable device data will be detailed in a separate document (that is, the digital biomarker analysis plan).

# 8. Unblinding Plan

The blinding/unblinding plan is in a separate document.

# 9. Change from the Protocol-Specified Statistical Analyses

For BOLD fMRI, the ***percent*** BOLD signal differences between views of highly palatable food (HF-HS and HF-HCCHO) versus non-food objects (for example, office supplies, furniture) was defined as the analysis variable. In data review in a blinded manner, quite a few extreme outliers were found, as there were a significant number of observations close to zero for non-food objects (the denominator in the percent change calculation). Therefore, the percent BOLD contrast was replaced with *absolute* BOLD contrast as the analysis variable.

# 10. References

Cappelleri JC, Bushmakin AG, Gerber RA, et al. Evaluating the Power of Food Scale in obese subjects and a general sample of individuals: development and measurement properties. *Int J Obes (Lond)*. 2009;33(8):913-922. https://doi.org/10.1038/ijo.2009.107

Cepeda-Benito A, Gleaves DH, Williams TL, Erath SA. The development and validation of the State and Trait Food-Cravings Questionnaires. *Behav Ther.* 2000;31(1):151-173. https://doi.org/10.1016/S0005-7894(00)80009-X

Flint A, Raben A, Blundell JE, Astrup A. 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

Lowe MR, Butryn ML, Didie ER, et al. The Power Food Scale. A new measure of the psychological influence of the food environment. *Appetite.* 2009;53(1):114-118. https://doi.org/10.1016/j.appet.2009.05.016

Moreno S, Rodríguez S, Fernandez MC, et al. Clinical validation of the trait and state versions of the Food Craving Questionnaire. *Assessment*. 2008;15(3):375-387. https://doi.org/10.1177/1073191107312651

Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt Impulsiveness Scale. *J Clin Psychol.* 1995;51(6):768-774. https://doi.org/10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1

Reid RC, Cyders MA, Moghaddam JF, Fong TW. Psychometric properties of the Barratt Impulsiveness Scale in patients with gambling disorders, hypersexuality, and methamphetamine dependence. *Addict Behav*. 2014;39(11):1640-1645. https://doi.org/10.1016/j.addbeh.2013.11.008

Stunkard AJ, Messick S. *Eating Inventory Manual*. Psychological Corporation, Harcourt Brace Jovanovich; 1988.

Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *J Psychosom Res*. 1985;29(1):71-83. https://doi.org/10.1016/0022-3999(85)90010-8

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

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

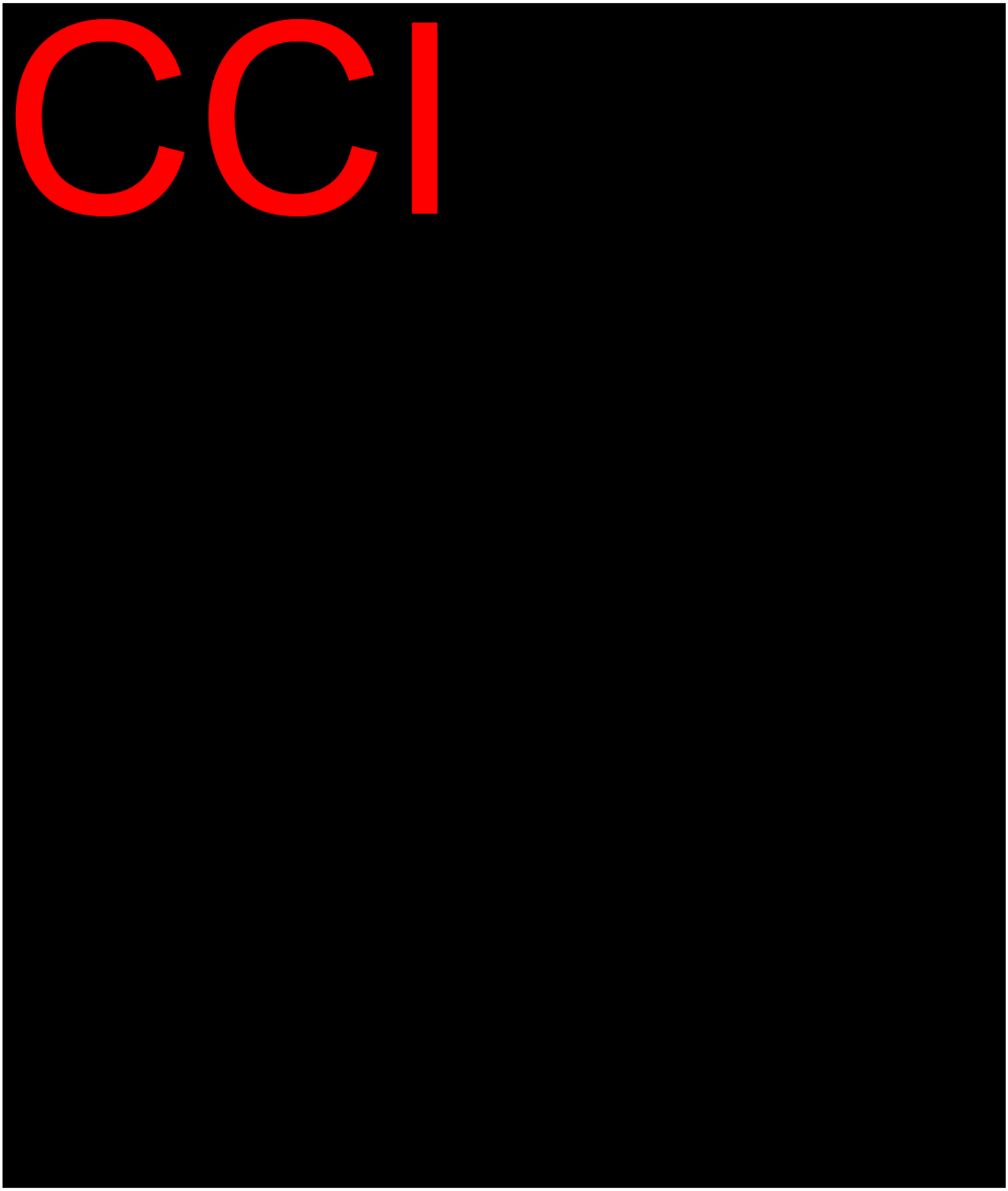
Womble LG, Wadden TA, Chandler JM, Martin AR. Agreement between weekly vs. daily assessment of appetite. *Appetite.* 2003;40(2):131-135. https://doi.org/10.1016/s0195-6663(02)00170-8

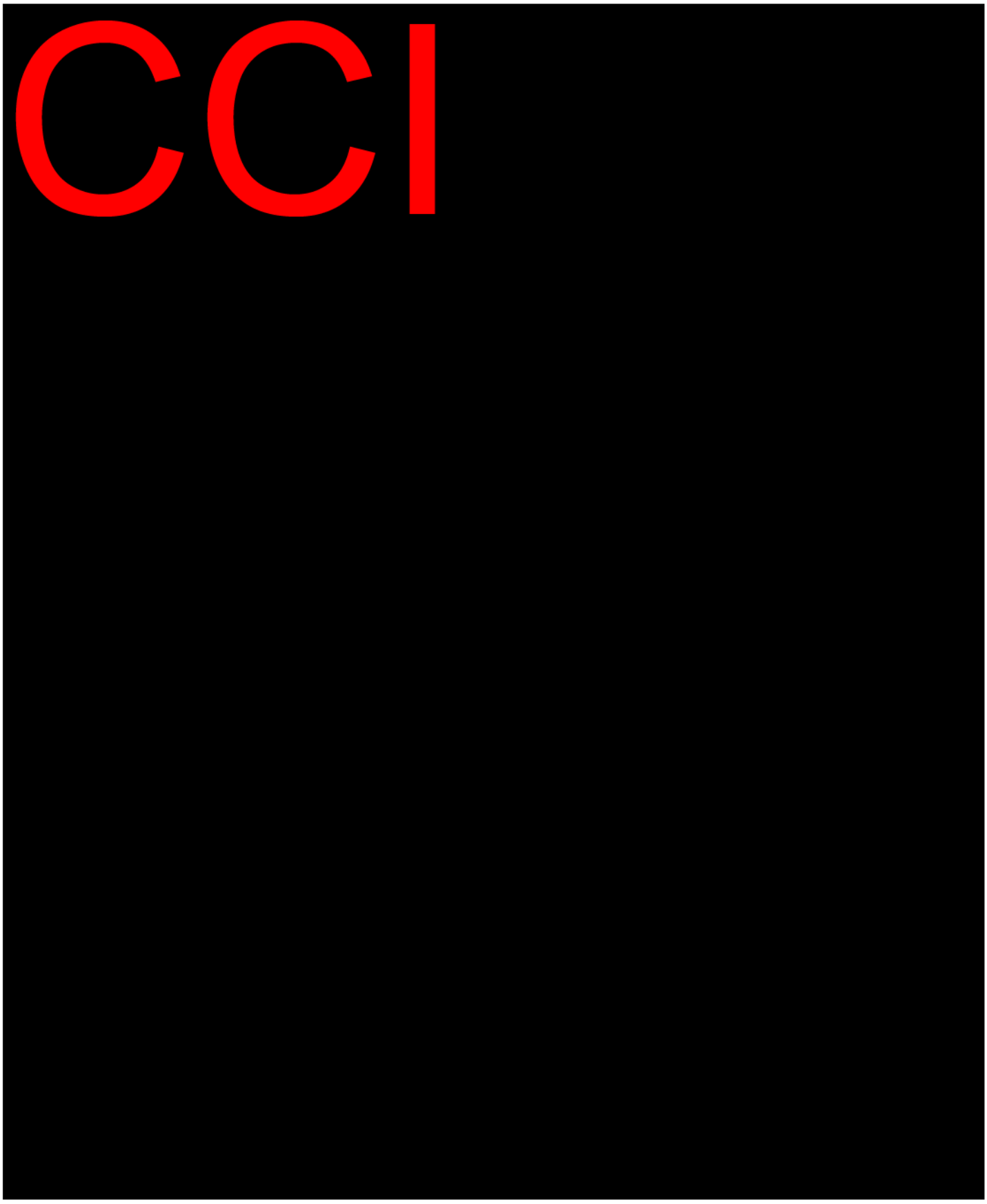
# 11. Appendices

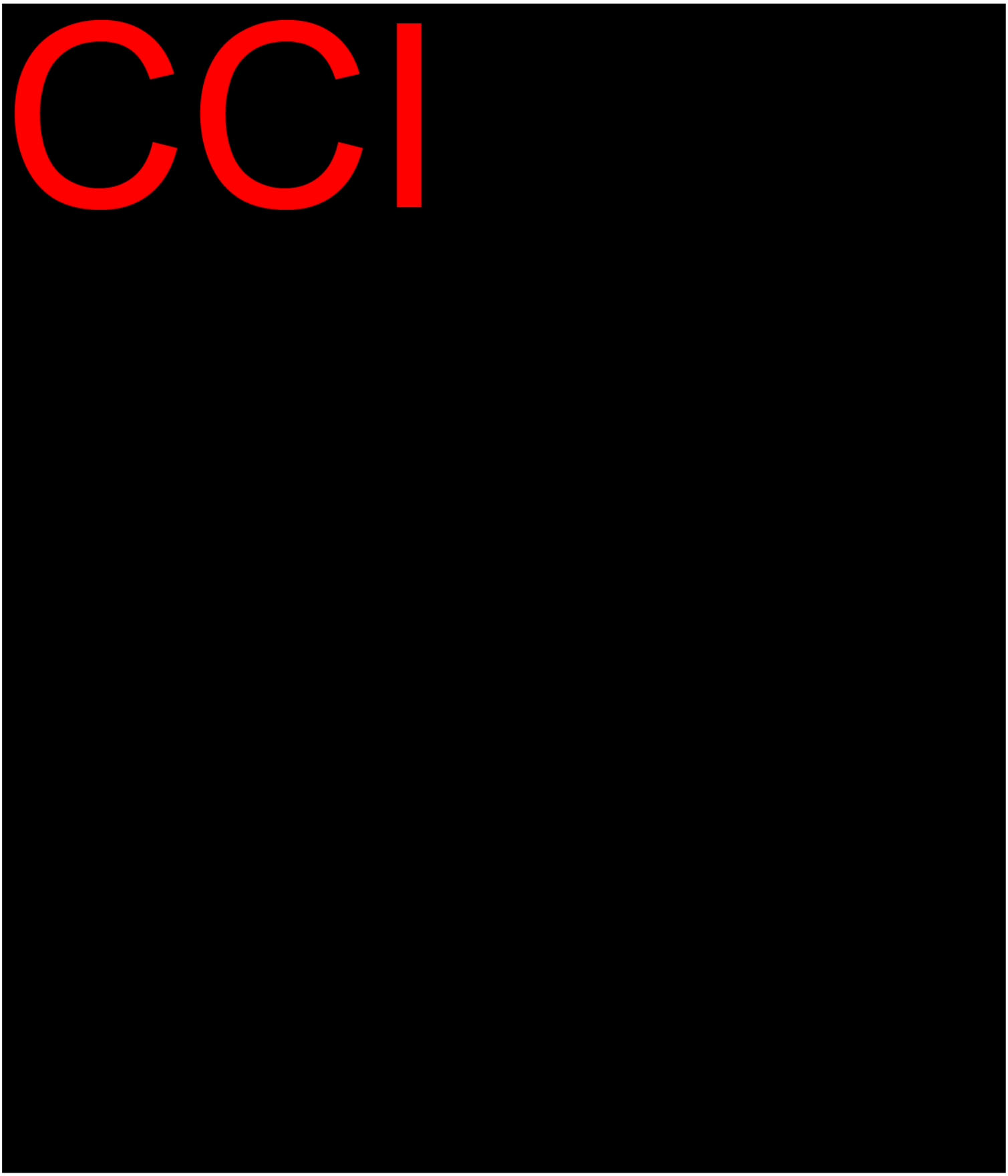
# Appendix 1

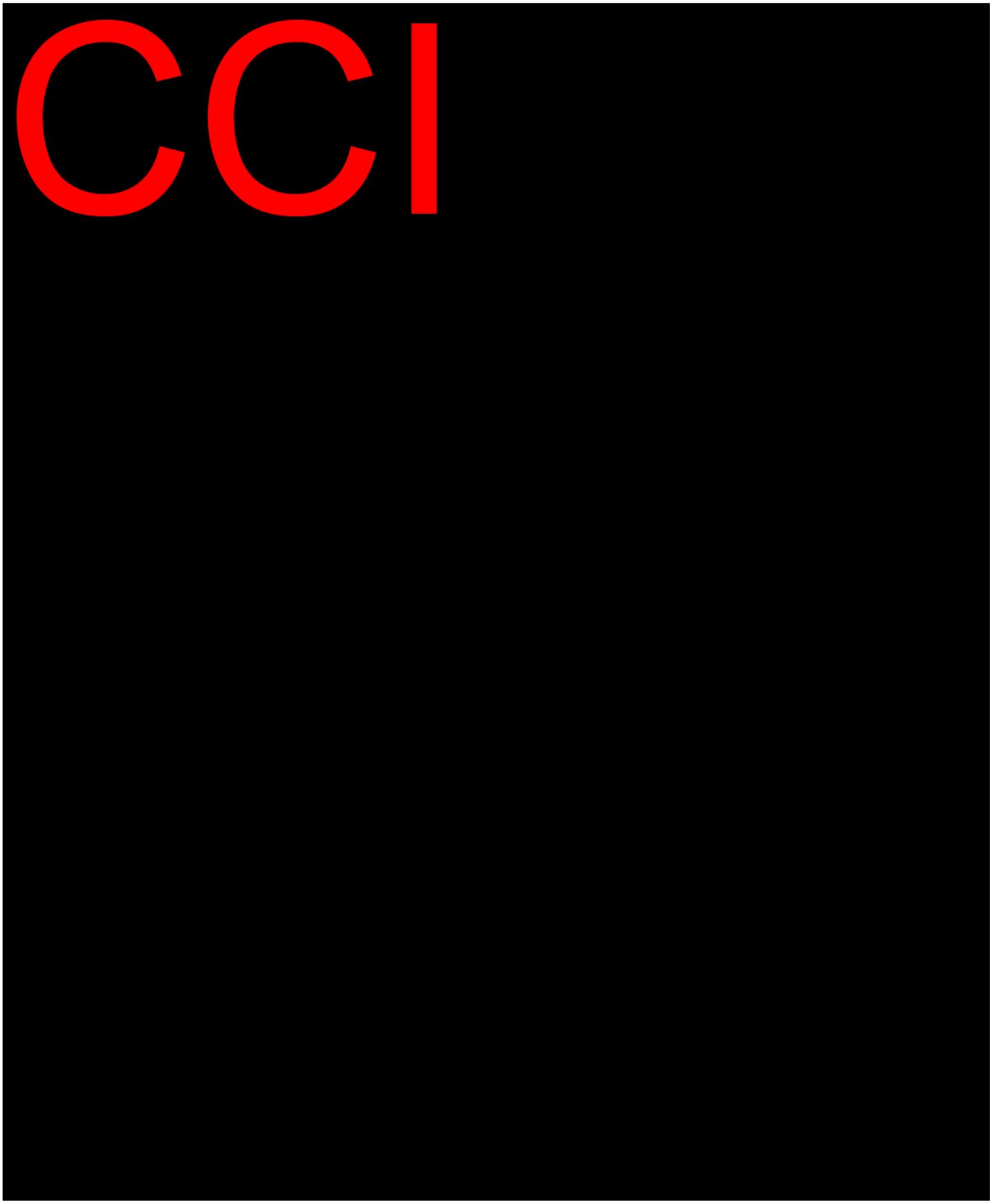
The search criteria for AEs of special safety topics and AESI are stored in the location below.

CLUWE:\\STATSCLSTR\LILLYCE\QA\LY3298176\common\AESI_Lab\Search criteria AESIs_TZP.xlsx


CCI


CCI


CCI


CCI

Signature Page for VV-CLIN-109361 v1.0

| | |
|---|---|
| Approval | PPD<br>20-Apr-2023 16:04:32 GMT+0000 |

Signature Page for VV-CLIN-109361 v1.0