Statistical Analysis Plan Version 1 -I8F-JE-GPHZ

Efficacy and Safety of Once-Weekly Tirzepatide in Participants with Obesity Disease: A Randomized, Double-Blind, Placebo-Controlled Trial (SURMOUNT-J)

NCT04844918

Approval Date: 11-Jul-2023

Title Page

Protocol Title: Efficacy and Safety of Once-Weekly Tirzepatide in Participants with Obesity Disease: A Randomized, Double-Blind, Placebo-Controlled Trial (SURMOUNT-J)

Protocol Number: I8F-JE-GPHZ

Compound Number: LY3298176

Short Title: Efficacy and Safety of Tirzepatide Once Weekly Versus Placebo in Participants with Obesity Disease (SURMOUNT-J)

Acronym: SURMOUNT-J

Sponsor Name: Eli Lilly Japan K.K.

Legal Registered Address: LILLY PLAZA ONE BLDG, 5-1-28 Isogamidori, Chuo-ku Kobe City, Hyogo 651-0086 Japan

Regulatory Agency Identifier Number(s): Not applicable

Table of Contents

Abbreviations and Terms

Term	Definition
ADA	anti-drug antibodies
AE	adverse event
AESI	adverse events of special interest
AHM	antihyperglycemic mediation
ALP	alkaline phosphatase
ALT	alanine transaminase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BG	blood glucose
BMI	body mass index
bpm	beats per minute
CEC	clinical endpoint committee
CI	confidence interval
CN	conventional
COVID-19	coronavirus disease 2019
CRF	case report form
CSR	clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
DBP	diastolic blood pressure
EAS	Efficacy Analysis Set
ECG	electrocardiogram
eCRF	electronic case report form
ED	early discontinuation
eGFR	estimated glomerular filtration rate
EQ-5D-5L	EuroQol - 5 Dimension - 5 Level

Term	Definition
EWL	excess weight loss
FAS	Full Analysis Set
FDA	Food and Drug Administration
GI	gastrointestinal
GIP	glucose-dependent insulinotropic polypeptide
GIPR	glucose-dependent insulinotropic polypeptide receptor
GLP-1	glucagon-like peptide-1
GLP-1R	glucagon-like peptide-1 receptor
HbA1c	hemoglobin A1c
HDL-C	high-density lipoprotein cholesterol
HFF	hepatic fat fraction
HLT	High Level Term
ICE	intercurrent event
ICH	International Council for Harmonisation
IGT	impaired glucose tolerance
ISR	injection-site reaction
IWQOL-Lite-CT	Impact of Weight on Quality of Life-Lite-Clinical Trials Version
JASSO	Japan Society for the Study of Obesity
LDL-C	low density lipoprotein cholesterol
Lilly	Eli Lilly and Company
LLT	Lowest Level Term
LY	LY3298176 (tirzepatide)
MACE	major adverse cardiovascular events
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intention-to-treat
MMRM	mixed model for repeated measures

Term	Definition
MRD	minimum required dilution
MRI	magnetic resonance imaging
NAb	neutralizing antibodies
NAFLD	non-alcoholic fatty liver disease
NASH	nonalcoholic steatohepatitis
nGIP	native glucose-dependent insulinotropic polypeptide
nGLP-1	native glucagon-like peptide-1
OGTT	oral glucose tolerance test
ORHP	obesity related health problem
PD	pharmacoydynamic
PDFF	proton density fat fraction
PG	plasma glucose
PGIS	Patient Global Impression of Status for Physical Activity
PHQ-9	Patient Health Questionnaire-9
PK	pharmacokinetic
PT	Preferred Term
QTcF	Fredericia's corrected QT interval
QW	once weekly
REML	restricted maximum likelihood
SAE	serious adverse event
SAP	statistical analysis plan
SAT	subcutaneous adipose tissue
SBP	systolic blood pressure
SD	standard deviation
SF-36v2	Short Form 36-item Health Survey Version 2
SI	Système International

Term	Definition
SMQ	Standardised MedDRA Query
SOC	System Organ Class
SS	Safety Analysis Set
T2DM	type 2 diabetes mellitus
TC	total cholesterol
TBL	total bilirubin
TE	treatment-emergent
TEAE	treatment-emergent adverse event
TG	triglyceride
UACR	urine albumin-to-creatinine ratio
ULN	upper limit of normal
VAS	Visual Analog Scale
VAT	visceral adipose tissue
VLDL-C	very low density lipoprotein cholesterol

Version history

This Statistical Analysis Plan (SAP) for study GPHZ is based on the protocol (b) dated 11JUN2021.

Table 1. SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1		Not Applicable	Original version

1. Introduction

Changes to the protocol-planned analyses are described in Section 4.9.

1.1. Objectives, Endpoints, and Estimands

1.1.1. Objectives and Endpoints

Table GPHZ.1.1. Objectives and Endpoints

Objectives	Endpoints
Primary	
To demonstrate that once-weekly tirzepatide, at doses of 10 mg and/or 15 mg, is superior to placebo in terms of body weight reduction from BL to 72 weeks in Japanese participants with obesity disease	• Mean percent change in body weight at Week 72 and • Proportion of participants who achieve ≥5% body weight reduction at Week 72
Secondary	
To compare the efficacy of once-weekly tirzepatide 10 mg and/or 15 mg with placebo for	
• Improvement in obesity-related health problems (72 weeks)	• Proportion of participants who met the following criteria at 72 weeks from BL o Improvement in at least 2 obesity-related health problems for participants with BL BMI ≥27 and <35 kg/m^2, or o Improvement in at least 1 obesity-related health problem for participants with BL BMI ≥35 kg/m^2
• Improvement in laboratory tests for obesity-related health problems (72 weeks)	• Mean changes in the following laboratory tests at 72 weeks from BL o OGTT 2-hr glucose (only for participants with IGT at BL) o Fasting glucose (only for participants with IGT at BL) o Fasting lipids (TG) (only for participants with hyperlipidemia at BL) o HFF (only for participants diagnosed as NAFLD by MRI at BL)
• Improvement of IGT (72 weeks)	• Proportion of participants who achieved improvements of IGT (only for participants with IGT at BL)
• Improvement of hyperlipidemia (72 weeks)	• Proportion of participants who achieved improvements of hyperlipidemia (only for participants with hyperlipidemia at BL)
• Improvement of NAFLD (72 weeks)[a]	• Proportion of participants who achieved improvements of NAFLD (only for participants who are diagnosed as NAFLD by MRI at BL)

Objectives	Endpoints
• **Body mass related parameters** o Change in body weight (24, 52, and 72 weeks)	• Proportion of participants who achieve the following criteria from BL o ≥5% body weight reduction (only at 24 and 52 weeks) o ≥7% body weight reduction o ≥10% body weight reduction o ≥15% body weight reduction o ≥20% body weight reduction • Mean change from BL in o Absolute body weight o BMI o Mean percent change from BL in body weight (only at 24 and 52 weeks) • Mean EWL
o Change in VAT and SAT (72 weeks)	• Mean change from BL in o VAT area (cm^2) o SAT area (cm^2) o VAT/SAT ratio • Mean percent change from BL in o VAT area (cm^2) o SAT area (cm^2) o VAT/SAT ratio • Proportion of participants who achieve VAT area <100 cm^2 (only for participants with VAT area≥100 cm^2 at BL) from BL
o Change in waist circumference (24, 52, and 72 weeks)	• Mean change from BL in waist circumference
• **IGT related parameters (only for participants with IGT at BL)** o Change in IGT related parameters (24, 52, and 72 weeks)	• Mean change from BL in o Fasting glucose o HbA1c o Fasting insulin o C-peptide o HOMA2-%B o HOMA2-%S
• **Hyperlipidemia related parameters (only for participants with hyperlipidemia at BL)** o Change in fasting lipids (24, 52, and 72 weeks)	Mean change from BL in o TG o TC o VLDL-C o LDL-C (direct) o HDL-C o Non-HDL-C o FFA

Objectives	Endpoints

• **NAFLD related parameters**
  ○ Change in NAFLD/NASH biomarkers (24, 52, and 72 weeks)
	• Mean change from BL in
	  ○ ALT
	  ○ AST
	  ○ GGT
	  ○ AST/ALT ratio
	  ○ K-18 (M30)
	  ○ Pro C3
	  ○ ELF
	  ○ FIB-4 index

• **Other parameters**
  ○ Change in blood pressure (24, 52, and 72 weeks)
	• Mean changes from BL in
	  ○ Systolic blood pressure
	  ○ Diastolic blood pressure

○ Change in uric acid (24, 52, and 72 weeks)	• Mean change from BL in uric acid
• **Patient-reported outcomes** ○ Change in patient-reported physical functioning (72 weeks)	• Mean change from BL in SF-36v2 acute form physical functioning domain score from BL • Mean change in IWQOL-Lite-CT Physical Function score

○ Change in patient-reported health status (72 weeks)	• Mean change in EQ-5D-5L score from BL
	  ○ Index score
	  ○ VAS score

○ Change in dietary evaluation based on the meal record (24, 52, and 72 weeks)[b]	• Mean change from BL in
	  ○ Total fats
	  ○ Total carbohydrates
	  ○ Total proteins
	  ○ Total caloric intake

○ Change in dietary evaluation based on appetite sensation (24, 52, and 72 weeks)[c]	• Mean change (VAS) from BL in
	  ○ Satiety
	  ○ Fullness
	  ○ Prospective food consumption
	  ○ Hunger
	  ○ Overall appetite score

To characterize the population PK of tirzepatide and explore the relationships between the tirzepatide concentration and efficacy, safety, tolerability, and pharmacogenetics measures	• Population PK and PD parameters

Objectives	Endpoints
Exploratory	
To compare the efficacy of once-weekly tirzepatide 10 mg and/or 15 mg with placebo for	
○ Improvement in obesity-related health problems at 72 weeks	○ Proportion of participants who met the following criteria at 72 weeks from BL ○ improvement in ≥2 obesity-related health problems for participants with BL BMI ≥27 and <35 kg/m^2 ○ improvement in ≥1 obesity-related health problems for participants with BL BMI ≥35 kg/m^2 ○ Proportion of participants who achieved the following number of improvements of obesity-related health problems regardless of BMI at BL ○ Achieved 2 or 3 improvements (only for participants with 3 defined obesity-related health problems) ○ Achieved 2 improvements (only for participants with 2 defined obesity-related health problems) ○ Achieved 1 improvement (only for participants with 1 defined obesity-related health problems)
• **IGT related parameters (all participants)** ○ Change in OGTT (24, 52, and 72 weeks)	• Mean change from BL in ○ OGTT 2-hr glucose ○ fasting glucose ○ insulin ○ C-peptide
• **Hyperlipidemia related parameters (all participants)** ○ Change in fasting lipids (24, 52, and 72 weeks)	• Mean change from BL in ○ TG ○ TC ○ VLDL-C ○ LDL-C (direct) ○ HDL-C ○ non-HDL-C ○ FFA

Objectives	Endpoints
• **Other parameters**	
○ Change in laboratory tests (24, 52, and 72 weeks)	• Mean change from BL in ○ Adiponectin ○ Leptin ○ hs-CRP ○ TNFα • Mean change from BL in ○ eGFR ○ Urinary albumin/creatinine ratio
• **Patient-reported outcomes**	
○ Change in SF-36v2 acute scores other than physical functioning (72 weeks)	• Mean change from BL in ○ role-physical score ○ bodily pain score ○ general health score ○ vitality score ○ social functioning score ○ role-emotional score ○ mental health score ○ physical component summary ○ mental component summary
○ Change in IWQOL-Lite-CT scores other than physical functioning (72 weeks)	• Mean change from BL in ○ physical score ○ psychosocial score ○ total score
○ Change in Patient Global Impression of Status (72 weeks)	• Shift of the ordinal response category from BL
To compare once-weekly tirzepatide (all doses combined) with placebo for participants with IGT at BL • Delayed progression to T2DM at 72 weeks	• Time to onset of T2DM during 72-week treatment period from BL
To explore once-weekly tirzepatide (all participants combined) 10 mg and/or 15 mg and placebo for • Immunogenicity assessment	• The frequency and percentage of participants with preexisting ADA, with TE ADA, and with neutralizing TE ADA to tirzepatide

Abbreviations: ADA = anti-drug antibodies; ALT = alanine aminotransferase; AST = aspartate aminotransferase; BL = baseline; BMI = body mass index; eGFR = estimated glomerular filtration rate; ELF = enhanced liver fibrosis; EWL = excess weight loss; FFA = free fatty acid; FIB-4 = Fibrosis-4; GGT = gamma-glutamyl transferase; HbA1c = glycated hemoglobin; HDL-C = high-density lipoprotein cholesterol HFF = hepatic fat fraction; HOMA2-%B = homeostasis model assessment estimates steady state beta cell function; HOMA2-%S = homeostasis model assessment estimates steady state insulin sensitivity; IGT = impaired glucose tolerance; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite-Clinical Trials Version; K-18 (M30) = keratin-18 M30 fragment; LDL-C = low-density lipoprotein cholesterol; NAFLD = non-alcoholic fatty liver disease; NASH = nonalcoholic steatohepatitis; OGTT = oral glucose tolerance test; PD = pharmacodynamic(s); PK = pharmacokinetic(s); Pro-C3 = a fragment of the NH2-terminal propeptide of type III procollagen; SAT = subcutaneous adipose tissue; SF-36v2 = Short Form 36-item Health Survey Version

2; T2DM = type 2 diabetes mellitus; TC = total cholesterol; TE ADA = treatment-emergent antidrug antibodies; TG = triglycerides; TNFα = tumor necrosis factor alpha; VAS = Visual Analog Scale; VAT = visceral adipose tissue; VLDL = very low-density lipoproteins.

[a] To compare the efficacy of once-weekly tirzepatide with placebo, improvement of NAFLD at 72 weeks will be combined with all doses.

[b] Individual daily meal intake information (photographs and record sheet), including breakfast, lunch, dinner, and snacks, will be collected on 3 consecutive or non-consecutive days during the week before the visit. These records will be collected only from participants who agree to participate in taking these measurements.

[c] Appetite sensation will be measured by a 100-mm VAS for parameters of hunger, fullness, satiety, and prospective food consumption, including snacks before and after meal recorded on 3 consecutive or non-consecutive days during the week immediately before the visit. These records will be collected only from participants who agree to participate in taking these measurements.

1.1.2. Estimands

The primary clinical question of interest is: What is the intervention difference between tirzepatide 10 mg and/or 15 mg QW and placebo, when used in conjunction with a reduced-calorie diet and increased physical activity for weight management, in mean percent change in body weight, and proportion of participants who achieve ≥5% body weight reduction, from randomization to 72 weeks, in obesity disease participants with BMI ≥27 kg/m^2 and at least 2 obesity-related health problems, or participants with BMI ≥35 kg/m^2 and at least 1 obesity-related health problem? The obesity-related health problems are defined as IGT, hyperlipidemia, and NAFLD.

1.1.2.1. Primary estimand

The estimand is described by the following attributes:

- Population: all randomized participants who received at least 1 dose of treatment.
- Endpoint: mean percent change from randomization to Week 72 in body weight, AND percentage of study participants who achieve ≥5% body weight reduction from randomization to Week 72.
- Treatment condition of interest: tirzepatide 10 mg and/or 15 mg QW vs. placebo, **excluding data after discontinuation of study drug**. Temporary interruptions and modification prior to discontinuation will be allowed as part of the treatment conditions.
- Handling of intercurrent events: the intercurrent events of treatment discontinuation for any reason is addressed by the treatment condition of interest attribute and handled by the hypothetical strategy.
- Population-level summary: difference in mean percent changes between treatment conditions, AND difference in response percentage between treatment conditions.

Rationale for estimand: the estimand aims to reflect treatment efficacy in an envisaged scenario in which the intercurrent events leading to treatment discontinuation would not occur.

This estimand is referred to as ‘efficacy’ estimand in the latter of this document.

1.1.2.2. Supplemental estimand(s)

The estimand is described by the following attributes:

- Population: all randomized participants who received at least 1 dose of treatment.
- Endpoint: mean percent change, from randomization to Week 72, in body weight, AND percentage of study participants who achieve ≥5% body weight reduction from randomization to Week 72.
- Treatment condition of interest: tirzepatide 10 mg and/or 15 mg vs. placebo, **regardless of study drug adherence**.
- Handling of intercurrent events: the intercurrent events of treatment discontinuation for any reason is addressed by the treatment condition of interest attribute, and handled by treatment policy strategy. Further details can be found in Section 4.3.3.1.
- Population-level summary: difference in mean percent changes between treatment conditions, AND difference in response percentage between treatment conditions.

Rationale for estimand: This aims at reflecting how participants with obesity disease with obesity-related health problems, are treated in clinical practice, and take into account both safety and efficacy.

This *de facto estimand* is referred to as the 'treatment-regimen' estimand in the latter sections of this document.

1.2. Study Design

Study GPHZ is a Phase 3, multicenter, randomized, placebo-controlled, double-blinded study to evaluate the safety and efficacy of tirzepatide 10 mg and 15 mg administered subcutaneously once weekly compared with placebo when used in conjunction with a reduced-calorie diet and increased physical activity for weight management, in participants who have

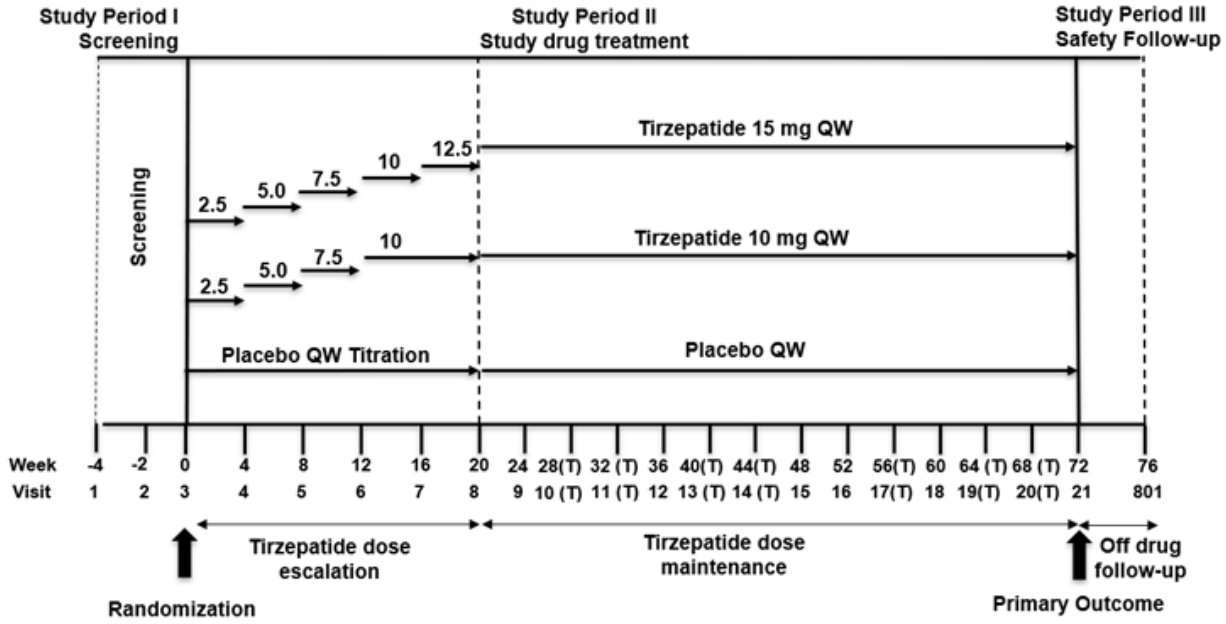
- A BMI of 27 kg/m^2 or greater and less than 35 kg/m^2 and at least 2 obesity-related health problems; OR
- A BMI of 35 kg/m^2 or greater and at least 1 obesity-related health problem.

Obesity-related health problems are defined as IGT, hyperlipidemia, or NAFLD.

All participants will undergo a 4-week screening period and a 72-week treatment period including a 20-week dose escalation. The safety follow-up period will be 4 weeks.

Study participants will be randomized in a 1:1:1 ratio to tirzepatide 10 mg QW, tirzepatide 15 mg QW, or placebo QW, stratified by

- baseline IGT (yes or no)
- baseline hyperlipidemia (yes or no)
- baseline NAFLD (yes or no), and
- sex (male, female).



Abbreviations: QW= once weekly; T= telephone visit.
All participants will be randomized to 72 weeks of treatment.

Figure 1.1. Study design for Clinical Protocol I8F-JE-GPHZ.

2. Statistical Hypotheses

The alternative hypotheses for the primary objective are the following:

- H_{10}: QW tirzepatide 10 mg is superior to placebo for percent change in body weight from randomization AND percentage of participants who achieve ≥5% body weight reduction at 72 weeks.
- H_{15}: QW tirzepatide 15 mg is superior to placebo for percent change in body weight from randomization AND percentage of participants who achieve ≥5% body weight reduction at 72 weeks.

The above 2 hypotheses will be tested in parallel, each at a 2-sided significance level of 0.025.

2.1. Multiplicity Adjustment

The type 1 error will be controlled in the strong sense at a two-sided significance level of 0.05 because two hypotheses (H_{10} and H_{15}) will be tested in parallel each at a 2-sided significance level of 0.025. Since superiority of each one of two hypotheses will be declared if and only if both statistical tests (regarding percent change in body weight and percentage of participants who achieve ≥5% body weight reduction) are significant, no multiplicity adjustment is planned for these 2 tests.

3. Analysis Sets

For purposes of analysis, the following populations are defined:

Table GPHZ.3.1. Description of Analysis Populations

Population	Description
Entered	All participants who sign informed consent form (ICF)
Randomized	All participants who are randomly assigned a study drug.
Modified Intent-to-Treat (mITT)	All randomly assigned participants who are exposed to at least 1 dose of study drug. Participants will be included in the treatment group to which they were randomized.

Table GPHZ.3.2. Description of Analysis Datasets

Analysis Datasets	Description
Efficacy Analysis Set (EAS)	Data obtained during the treatment period from the modified intent-to-treat (mITT) population, excluding data after discontinuation of study drug (last dose date + 7 days) Regarding imaging measurements (MRI and CT), data obtained during the treatment period from the mITT population, excluding data after discontinuation of study drug (last dose date + 14 days)
Full Analysis Set (FAS)	Data obtained during treatment period from the mITT population, regardless of adherence to study drug.
Safety Analysis Set (SS)	Data obtained during the treatment plus follow up period from the mITT population, regardless of adherence to study drug.

Unless otherwise specified, for analyses guided by the "efficacy" estimand, EAS will be used. For analyses guided by the "treatment-regimen" estimand, FAS will be used. For safety analysis, SS will be used.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the SAP or the clinical study report. Additional exploratory data analyses may be conducted as deemed appropriate.

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and the confidence interval will be calculated at 95%, 2-sided. In statistical summaries and analyses, data will be analyzed as randomized.

Unless specified otherwise, efficacy and safety will be assessed using the modified intent-to-treat (mITT) population. Baseline is defined as the last non-missing data collected at randomization (prior to first dosing of study drug). Unless specified otherwise, safety assessments will be guided by an estimand comparing safety of tirzepatide doses with placebo irrespective of adherence to study drug. Thus, safety analysis will be conducted using SS.

Summary statistics for continuous measures will include sample size, mean, SD, median, minimum, and maximum. The analysis model to make comparisons among treatment groups relative to continuous measurements assessed over time will be a mixed model for repeated measures (MMRM), with terms of treatment, visit, and treatment by visit interaction, stratification factors, and baseline measurement as a covariate.

Kaplan Meier method will be used for estimation of cumulative event free survival rates over time, and Cox proportional hazards regression analysis will be used to compare hazard rates among treatments.

Summary statistics for categorical measures (including categorized continuous measures) will include sample size, frequency, and percentages. Logistic regression will be used to examine the treatment difference in binary efficacy outcomes if there is a need to adjust for covariates. Otherwise, Fisher's exact test will be used to examine the treatment difference in categorical outcomes.

Summary statistics for discrete count measures will include sample size, mean, SD, median, minimum, and maximum.

For the safety related parameters, the definition of baseline and postbaseline are specified in Table GPHZ.4.1.

Table GPHZ.4.1. Baseline and Postbaseline Definitions for Safety Analysis

Analysis Dataset	Analysis Type	Baseline	Postbaseline
SS	1.1) Treatment-Emergent Adverse Events	The baseline period is defined as the start of screening and ends prior to the first dose in of study treatment (typically at Week 0).	Starts after the first dose of study treatment and ends at the end of the study period (including off-drug follow up visit).
SS	1.2) Treatment-Emergent Abnormal Labs[a], Vital Signs, and ECGs.	Baseline will include all scheduled and unscheduled measurements during the baseline period (Visit 1 to Visit 3)	Postbaseline will be defined as above (1.1). All scheduled and unscheduled measurements will be included.
SS	1.3) Change from Last Baseline to Week *xx* and to Last Postbaseline for Labs[a],Vital Signs, and ECGs.	The last scheduled and unscheduled non-missing assessment recorded during the baseline period defined above (1.2).	Postbaseline will be defined as above (1.1). Only scheduled visits will be included. The early discontinuation (ED) visits are considered scheduled visits.

Abbreviations: ECGs = electrocardiogram; SS = Safety Analysis Set.

[a] Immunogenicity related analysis is specified in Section 4.5.5.5.

For the primary and secondary efficacy endpoint analyses, data for participants with missing values at the 72-week visit will be imputed based on the method described in Section 4.3.2. Otherwise, missing values will not be explicitly imputed except for the parameters with only 1 postbaseline measure during the analysis period per schedule of activity, where last observation carried forward (LOCF) approach will be applied to impute the endpoint when ED measure is available.

End of study participation for a participant will be the earliest of date of death, date of withdrawal from further participation in the study, or date of safety follow-up visit (Visit 801). For participants considered to be lost-to-follow-up, end of study participation will be the date of lost-to-follow-up reported by the investigator. Participant data included in the database after the last date of study participation (including safety follow-up period) will be excluded from statistical analysis.

Statistical treatment comparisons will only be performed between tirzepatide 10 mg or 15 mg, and placebo. Since the trial is not adequately powered to detect difference among tirzepatide doses, comparisons among tirzepatide doses will not be performed unless otherwise specified.

Statistical summaries and results of statistical analyses will be displayed in the following order: Placebo, tirzepatide 10 mg, tirzepatide 15 mg, and pooled tirzepatide (all doses combined, if necessary).

Not all analyses described in this SAP will necessarily be included in the Clinical Study Reports (CSRs). Any analysis described in this SAP and not provided in the CSR would be available upon request. Not all displays will necessarily be created as a “static” display. Some may be incorporated into interactive display tools instead of or in addition to a static display.

4.2. Participant Dispositions

Reasons for screen failure as reported by investigators will be summarized.

Frequency counts and percentages of all participants screened, randomized, and receiving at least 1 dose of study drug will be summarized by treatment group, if applicable.

Of the randomized population, frequency counts and percentages of participants who completed the study, prematurely discontinued the study (and/or study drug), including reason for premature discontinuation, will be summarized by treatment group.

A Kaplan-Meier analysis of time from randomization to premature discontinuation from study and/or study treatment by treatment group will be provided. A time-to-event analysis of premature study treatment discontinuation will also be conducted.

4.3. Primary Endpoints Analysis

The assessment of efficacy objectives will be guided by the "efficacy" estimand using the EAS.

4.3.1. Definition of endpoints

The primary efficacy measure will be percent change in body weight AND percentage of participants who achieve ≥5% body weight reduction from randomization at 72 weeks. The percent change in body weight at each nominal visit is defined as:

*(postbaseline body weight [kg] – baseline body weight [kg]) / baseline body weight [kg] * 100*.

Both percent change in body weight and percentage of participants who achieve ≥5% body weight reduction will be summarized by treatment and nominal visit (week) from randomization to 72 weeks.

4.3.2. Main analytical approach

The analysis related to efficacy estimand will be conducted utilizing data in the EAS.

4.3.2.1. Main Analytical Approach for Continuous Measures

For the mean percent change in body weight from randomization, a MMRM will be conducted. Restricted maximum likelihood (REML) will be used to obtain model parameter estimates and the Kenward-Roger option will be used to estimate the denominator degrees of freedom. The response variable of MMRM will be the percent change in body weight from baseline values, obtained at each scheduled post-baseline visit.

For MMRM, the independent variables of analysis model are treatment group, visit, and treatment-by-visit interaction, stratification factors: baseline IGT (yes or no); baseline hyperlipidemia (based on TG; yes or no); baseline NAFLD (yes or no) and sex (female, male) as fixed effects, and baseline body weight as a covariate. An unstructured covariance structure will model relationship of within-patient 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
- compound symmetry

The first covariance structure that converges will be used.

With the aid of the MMRM analysis, 2-sided 95% CI for mean percent change in body weight from randomization to the 72-week visit for tirzepatide 10 mg or 15 mg QW, compared to placebo, will be derived and summarized. The resulting least squares mean (LSM) estimates of mean percent change in body weight from baseline will be plotted by visit and by study treatment.

4.3.2.2. Main Analytical Approach for Categorical Measures

For the percentage of participants achieving at least 5% body weight reduction from randomization over time, a logistic regression model will be used with the response variable of the percentage of participants achieving at least 5% body weight reduction at each scheduled postbaseline visit.

A logistic regression model with terms of treatment group stratification factors: baseline IGT (yes or no); baseline hyperlipidemia (based on TG; yes or no); baseline NAFLD (yes or no) and sex (female, male) as fixed effects, and baseline body weight as a covariate, will be conducted for percentage of participants achieving at least 5% body weight reduction from randomization at the 72 week visit. Missing body weight measurement at 72 week will be imputed by the predicted value from MMRM model aforementioned, then the continuous measurements will be dichotomized into status of achieving at least 5% body weight reduction (Yes or No). A logistic regression will be utilized to analyze proportion of participants with at least 5% body weight reduction at each nominal visit from randomization through 72 weeks.

With the aid of the logistic regression model, 2-sided 95% CI and odds ratio for percentage of participants achieving at least 5% body weight reduction from baseline to the 72-week visit between tirzepatide 10 mg or 15 mg QW and placebo will be derived.

4.3.3. Supplementary analyses

4.3.3.1. Analysis Related to the Treatment Regimen Estimand

The analysis related to the “treatment-regimen” estimand will be conducted using data in the FAS.

The analysis for the mean percent change in body weight will be conducted utilizing ANCOVA. The response variable for the ANCOVA model will be percent change in body weight from randomization to Week 72. This model will include terms of baseline IGT (yes or no); baseline hyperlipidemia (based on TG; yes or no); baseline NAFLD (yes or no) and sex (female, male) as fixed effects and baseline body weight as a covariate. The ANCOVA analysis will be conducted with multiple imputation of missing body weight at Week 72 and statistical inference over multiple imputation of missing data guided by Rubin (1987).

With the aid of the ANCOVA model, 2-sided 95% CI for mean change in percent body weight from baseline to the 72-week visit between tirzepatide 10 mg or 15 mg QW and placebo will be derived.

With the aid of the logistic regression model, 2-sided 95% CI and odds ratio for percentage of participants achieving at least 5% body weight reduction from baseline to the 72-week visit between tirzepatide 10 mg or 15 mg QW and placebo will be derived.

4.3.3.2. Methods for Multiple Imputations

For efficacy analyses relative to the "treatment-regimen" estimand, the intercurrent events (ICEs) and the resulting missing values will be handled as follows:

- **Category 1:** for missing data solely due to an exceptional circumstances, such as a pandemic or natural disaster (after other reasons for missing data are ruled out), considers the missing data as missing at random. The missing data will be imputed using all nonmissing data of the primary outcome measurement from the same treatment arm.
- **Category 2:** for missing data due to all other reasons, it will be imputed based on retrieved dropouts in the same treatment arm, defined as observed primary outcome measurements, from participants in the same treatment group who had their efficacy assessed after early discontinuation of study drug. In cases where there are not enough retrieved dropouts to provide a reliable imputation model (for example, the model implemented by the SAS program does not converge), an alternative multiple imputation method with reference to the placebo group (that is, placebo multiple imputation) will be used.

4.4. Secondary Endpoint Analysis

4.4.1. Definition of endpoints

4.4.1.1. Definition of body mass related parameters

Body mass index (kg/m^2) will be derived using body weight in kilograms divided by the square height in meters.

Excess weight loss (%) will be derived using [(initial weight) – (current weight)]/[(initial weight) – (ideal weight)] x 100), with ideal weight considered as the weight corresponding to a BMI = 22 kg/m^2.

4.4.1.2. Defined Obesity-Related Health Problems

The definition of baseline obesity-related health problems (ORHPs) are:

- **IGT** from 2-hour OGTT results: Fasting glucose or 0-hour OGTT = 110-125 mg/dL (6.1-6.9 mmol/L) and/or 2-hour OGTT = 140-199 mg/dL (7.8-11.0 mmol/L).
- **Hyperlipidemia**: triglyceride (TG) ≥150 mg/dL at baseline
- **NAFLD**: HFF ≥5% by MRI at baseline

Definition of IGT is based on the GPHZ protocol (b). For each subjects, baseline ORHPs are determined by the measurements (not by the inputs of IWRS).

The definition of improvements of each defined obesity-related health problems are:

- **IGT** (only for participants with IGT at baseline)
    - Return to normal blood glucose range
        - OGTT 0-hour glucose <110 mg/dL AND

    - OGTT 2-hour glucose <140 mg/dL
- **Hyperlipidemia** (only for participants with hyper-triglyceridemia [triglyceride (TG) ≥150 mg/dL] at baseline)
    - Return to normal range (TG <150 mg/dL) OR
    - Reduced by 30% from baseline
- **NAFLD** (only for participants with diagnosis of NAFLD [HFF ≥5%] by MRI at baseline)
    - Return to normal range (HFF <5%) OR
    - Reduced by 30% from baseline

Incident diabetes is followed by Japanese Clinical Practice Guideline for Diabetes 2019 (Araki et al. 2020) and defined when any 1 of the following occur after randomization:

- unequivocal hyperglycemia (random glucose ≥200 mg/dL) with signs or symptoms of hyperglycemia (e.g., dry mouth, polyposia, polyuria, body weight loss, or diabetic retinopathy).
- within a 4 week period, the following laboratory evidence of diabetic type criteria including either HbA1c ≥6.5%, or glucose levels (fasting glucose or OGTT 0 hour glucose ≥126 mg/dL and/or OGTT 2 hour glucose ≥200 mg/dL) are observed:
    - Both glucose level and HbA1c meets diabetic type criteria
    - Only glucose level meets diabetic type criteria
        - With typical symptoms of DM or definite diabetic retinopathy
        - Without typical symptoms of DM or definite diabetic retinopathy, re-exam within 1 month and:
            - Both glucose level and HbA1c meets diabetic type criteria
            - Only glucose level meets diabetic type criteria
            - Only HbA1c meets diabetic type criteria
    - Only HbA1c meets diabetic type criteria, re exam within 1 month and
        - Both glucose level and HbA1c meet diabetic type criteria
        - Only glucose level meets diabetic type criteria
- initiation of any medication for the treatment of diabetes

4.4.1.3. Definition of composites of improvements of Obesity-Related Health Problems

Regarding the following endpoint in the secondary endpoint, definition of response is described below.

Proportion of participants who met the following criteria at 72 weeks from BL

- Improvement in at least 2 obesity related health problems for participants with BL BMI >27 and <35 kg/m^2, or
- Improvement in at least 1 obesity related health problem for participants with BL BMI >35 kg/m^2

Responder is defined as:

- Post-baseline number of ORHP is 1 or 0 for participants with BL BMI >27 and <35 kg/m^2
- Post-baseline number of ORHP is 0 for participants with BL BMI >35 kg/m^2

4.4.1.4. Definition of IGT related parameters

Homeostatic Model Assessment for Beta Cell and Insulin Sensitivity (HOMA2 %B and HOMA2 %S): Fasting glucose and insulin values will be used to calculate beta-cell function and insulin sensitivity using the updated Homeostasis Model Assessment (HOMA2) (Wallace et al. 2004).

4.4.1.5. Definition of parameters related to 2-hour OGTT

The following parameters will be collected during 2-hour OGTT for each scheduled OGTT visit:

- glucose
- insulin, and
- C-peptide.

For each above parameter, the mean concentration over 120 minutes plots will be created by glycemia status at randomization and by treatment group and by OGTT visit (that is, at baseline and at Week 72).

For each above parameter, the area under the curve from time zero to 2 hours $(AUC)_{(0\text{-}2h)}$ during OGTT will be calculated using the trapezoidal rule. The respective $AUC_{(0\text{-}2h)}$ will be summarized by treatment group and by OGTT visit.

In addition, the following OGTT derived parameters (Araki et al. 2020) will be calculated at each OGTT visit for each participant with OGTT:

- Glycemic response categories:
    - Normal:
        - OGTT 0 hour glucose <110 mg/dL **AND**
        - OGTT 2-hour glucose <140 mg/dL
    - Borderline type:
        - OGTT 0 hour glucose =110-125 mg/dL (6.1-6.9 mmol/L) **OR**
        - OGTT 2-hour glucose =140-199 mg/dL (7.8-11.0 mmol/L)
    - Suspected T2DM:
        - OGTT 0 hour glucose ≥126 mg/dL **AND**
        - OGTT 2-hour glucose ≥200 mg/dL

- Matsuda Index = $\frac{10000}{\sqrt{(FPG \times FPI) \times (\bar{G} \times \bar{I})}}$, where FPG is the fasting glucose, FPI is the fasting insulin, $\bar{G}$ denotes the mean glucose during OGTT, and $\bar{I}$ denotes the mean insulin during OGTT

- Insulin secretion = $\left(\frac{\Delta C_{pep}}{\Delta G}\right)_{0-120}$, where ΔC_{pep} denotes the incremental area under the plasma C-peptide curve during the 2-hour OGTT, and ΔG denotes the incremental area under the plasma glucose concentration curve during 2-hour OGTT
    - Insulin secretion index = $\frac{\left(\frac{\Delta C_{pep}}{\Delta G}\right)_{0-120}}{normal\ level\ of\ insulin\ secretion}$, where normal level of insulin secretion is the median value for insulin secretion, ie, $\left(\frac{\Delta C_{pep}}{\Delta G}\right)_{0-120}$, in participants with normal glucose tolerance at baseline
    - Normal insulin secretion: insulin secretion index >70%
    - Moderate impairment of insulin secretion: 50% to 70% (inclusive) of the insulin secretion index
    - Severe impairment of insulin secretion: <50% of the insulin secretion index
- β–cell function = Insulin secretion * Matsuda index = $\left(\frac{\Delta C_{pep}}{\Delta G}\right)_{0-120} * \frac{10000}{\sqrt{(FPG \times FPI) \times (\bar{G} \times \bar{I})}}$

4.4.2. Main analytical approach

4.4.2.1. Main Analytical Approach for Continuous Measures with Only 1 Postbaseline Measure

For mean change from randomization or actual measurement at 72 week visit, an ANCOVA model will be used.

An ANCOVA model with terms of treatment group, stratification factors: baseline IGT (yes or no); baseline hyperlipidemia (based on TG; yes or no); baseline NAFLD (yes or no) and sex (female, male) as fixed effects, baseline measurement as a covariate.

For measurement with only 1 postbaseline measure during the analysis period per schedule of activity, where LOCF approach will be applied to impute the endpoint when ED measure is available.

4.4.2.2. Main Analytical Approach for Categorical Measures with Only 1 Postbaseline Measure

For the percentage of participants achieving improvement criteria from randomization at 72 week visit, a logistic regression model will be used.

A logistic regression model with terms of treatment group, stratification factors: baseline IGT (yes or no); baseline hyperlipidemia (based on TG; yes or no); baseline NAFLD (yes or no) and sex (female, male) as fixed effects. Baseline measurement will be included as a covariate if appropriate.

For measurement with only 1 postbaseline measure during the analysis period per schedule of activity, where LOCF approach will be applied to impute the endpoint when ED measure is available.

4.4.2.3. Analytical Approach for Binary Outcome with Repeated Measures

For some binary outcomes, longitudinal logistic regression with repeated measurements will be used with the independent variables of analysis model are treatment group, visit, and treatment-by-visit interaction, stratification factors: baseline IGT (yes or no); baseline hyperlipidemia (based on TG; yes or no); baseline NAFLD (yes or no) and sex (female, male) as fixed effects, and baseline body weight as a covariate.

4.4.3. Secondary and Exploratory Endpoints

Table GPHZ.4.2. Secondary and Exploratory Efficacy Measures for Improvement of Obesity-related Health Problems

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Percentage of participants who met the following criteria at 72 weeks from BL	Improvement in at least 2 obesity related health problems for participants with BL BMI >27 and <35 kg/ m^2, or Improvement in at least 1 obesity related health problem for participants with BL BMI >35 kg/ m^2	Efficacy analysis set	Logistic model in Section 4.4.2.2	Definition of endpoint: 4.4.1.3
	Improvement in at least 2 obesity-related health problems	Efficacy analysis set - participants with BL BMI ≥27 and <35 kg/m^2	Logistic model in Section 4.4.2.2	Exploratory
	Improvement in at least 1 obesity-related health problem	Efficacy analysis set - participants with BL BMI ≥35 kg/m^2	Logistic model in Section 4.4.2.2	Exploratory
Improvement in laboratory tests for obesity-related health problems (72 weeks)	Mean changes in OGTT 2-hr glucose at 72 weeks from BL	Efficacy analysis set - participants with IGT at BL	MMRM model in Section 4.3.2.1	
	Mean changes in fasting glucose at 72 weeks from BL	Efficacy analysis set - participants with IGT at BL	MMRM model in Section 4.3.2.1	
	Mean changes in Fasting lipids (TG) at 72 weeks from BL	Efficacy analysis set - participants with hyperlipidemia at BL	MMRM model in Section 4.3.2.1	
	Mean changes in HFF at 72 weeks from BL	Efficacy analysis set - participants diagnosed as NAFLD by MRI at BL	ANCOVA model in Section 4.4.2.1	
Improvement of IGT (72 weeks)	Proportion of participants who achieved improvements of IGT	Efficacy analysis set - participants with IGT at BL	Logistic model in Section 4.3.2.2	

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Improvement of hyperlipidemia (72 weeks)	Proportion of participants who achieved improvements of hyperlipidemia	Efficacy analysis set - participants with hyperlipidemia at BL	Longitudinal logistic model in Section 4.4.2.3	
Improvement of NAFLD (72 weeks)	Proportion of participants who achieved improvements of NAFLD	Efficacy analysis set - participants diagnosed as NAFLD by MRI at BL	Logistic model in Section 4.3.2.2	Compare TZP 10 mg, 15 mg and pooled TZP doses with placebo
Percentage of participants who met the following criteria at 72 weeks from BL	Improvement in at 2 or 3 obesity related health problems	Efficacy analysis set - participants with 3 defined obesity-related health problems	Logistic model in Section 4.3.2.2	Exploratory
	Improvement in at 2 obesity related health problems	Efficacy analysis set - participants with 2 defined obesity-related health problems	Logistic model in Section 4.3.2.2	Exploratory
	Improvement in at 1 obesity related health problems	Efficacy analysis set - participants with 3 defined obesity-related health problems	Logistic model in Section 4.3.2.2	Exploratory

Abbreviations: ANCOVA = analysis of covariance; BL = baseline; BMI = body mass index; IGT = impaired glucose tolerance; MMRM = mixed model for repeated measures; NAFLD = non-alcoholic fatty liver disease; OGTT = oral glucose tolerance test; TG = triglycerides; TZP = tirzepatide.

Table GPHZ.4.3. Secondary and Exploratory Efficacy Measures for Body Mass

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Change in body weight (24, 52, and 72 weeks)	Proportion of participants who achieve the following criteria from BL • ≥5% body weight reduction (only at 24 and 52 weeks) • ≥7% body weight reduction • ≥10% body weight reduction • ≥15% body weight reduction • ≥20% body weight reduction	Efficacy analysis set	Logistic model in Section 4.3.2.2	
	Mean percent change from BL in • Body weight (%) (Only at 24 and 52 weeks) Mean change from BL in • Absolute body weight (kg) • BMI	Efficacy analysis set	MMRM model in Section 4.3.2.1	
EWL	EWL defined in Section 4.4.1.1	Efficacy analysis set	MMRM model in Section 4.3.2.1	
Change in VAT and SAT (72 weeks)	Mean change from BL in • VAT area (cm^2) • SAT area (cm^2) • VAT/SAT ratio	Efficacy analysis set	ANCOVA model in Section 4.4.2.1	
	Mean percent change from BL in • VAT area (cm^2) • SAT area (cm^2) • VAT/SAT ratio	Efficacy analysis set	ANCOVA model in Section 4.4.2.1	

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
	Proportion of participants who achieve VAT <100 cm^2	Efficacy analysis set - participants with VAT ≥100 cm^2 at BL	Logistic model in Section 4.4.2.2	
Change in waist circumference (24, 52, and 72 weeks)	Mean change from BL in waist circumference	Efficacy analysis set	MMRM model in Section 4.3.2.1	

Abbreviations: ANCOVA = analysis of covariance; BL = baseline; BMI = body mass index; EWL = excess weight loss; MMRM = mixed model for repeated measures; SAT = subcutaneous adipose tissue; VAT = visceral adipose tissue.

Table GPHZ.4.4. Secondary and Exploratory Efficacy Measures for IGT, Hyperlipidemia, NAFLD and Other Parameters

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Change in IGT related parameters (24, 52, and 72 weeks)	Mean change from BL in • Fasting glucose • HbA1c • Fasting insulin • C-peptide • HOMA2-%B • HOMA2-%S • Matsuda Index • Insulin Secretion • Beta-cell function	(1)Efficacy analysis set ; (2) Efficacy analysis set - participants with IGT at BL; (3) Efficacy analysis set - participants with normoglycemic at BL	MMRM model in Section 4.3.2.1	
Change in hyperlipidemia related parameters (24, 52, and 72 weeks)	Mean change from BL in • TG • TC • VLDL-C • LDL-C (direct) • HDL-C • Non-HDL-C • FFA	(1) Efficacy analysis set - participants with hyperlipidemia at BL; (2) Efficacy analysis set	MMRM model in Section 4.3.2.1	
Change in NAFLD related parameters (24, 52, and 72 weeks)	Mean change from BL in • ALT • AST • GGT • AST/ALT ratio • K-18 (M30) • Pro C3 • ELF • FIB-4 index	(1) Efficacy analysis set; (2) Efficacy analysis set - participants with NAFLD at BL	MMRM model in Section 4.3.2.1	

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Change in blood pressure (24, 52, and 72 weeks)	Mean change from BL in • Systolic blood pressure • Diastolic blood pressure	Efficacy analysis set	MMRM model in Section 4.3.2.1	
Change in uric acid (24, 52, and 72 weeks)	Mean change from BL in • Uric acid	Efficacy analysis set	MMRM model in Section 4.3.2.1	
Change in laboratory tests (24, 52, and 72 weeks)	Mean change from BL in • Adiponectin (high molecular weight and total) • Leptin • hs-CRP • TNFα • eGFR • Urinary albumin/creatinine ratio	Efficacy analysis set	MMRM model in Section 4.3.2.1	Exploratory objective
Time to onset of T2DM during 72 week treatment period	Onset of T2DM during 72 week treatment period from BL	Safety analysis set - participants with IGT at BL	Kaplan Meier plot will be used for estimation of survival rates over time Log-rank test will be used to compare survival curves	Exploratory objective To compare TZP 10 mg, 15mg and pooled TZP doses with placebo for participants with IGT at BL
Change in eGFR	Mean change from BL in • eGFR	Efficacy analysis set	MMRM model in Section 4.3.2.1	Exploratory objective added to protocol planned analyses

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Shift in eGFR	Shift table of minimum baseline to minimum postbaseline	Efficacy analysis set		Exploratory objective added to protocol planned analyses
Change in UACR	Mean change from BL in UACR	Efficacy analysis set	MMRM model in Section 4.3.2.1	Exploratory objective added to protocol planned analyses
Improvement of hyperuricemia (72 weeks)	Proportion of participants who achieved improvements of hyperuricemia. Definition of improvement of hyperuricemia: Uric acid 7 mg/dL	Efficacy analysis set – participants with hyperuricemia at BL defined as uric acid > 7 mg/dL	Longitudinal logistic model in Section 4.4.2.3	Exploratory objective added to protocol planned analyses
Improvement of hypertension (72 weeks)	Proportion of participants who achieved improvements of hypertension. Definition of improvement of hypertension: Diastolic blood pressure <80 mmHg or Systolic blood pressure <130 mmHg.	Efficacy analysis set – participants with hypertension at BL defined as diastolic blood pressure ≥80 mmHg or Systolic blood pressure ≥130 mmHg.	Logistic model in Section 4.4.2.2	Exploratory objective added to protocol planned analyses
Change in NAFLD related parameters (24, 52, and 72 weeks)	Mean change from BL in • NAFLD fibrosis score	(1) Efficacy analysis set; (2) Efficacy analysis set - participants with NAFLD at BL	MMRM model in Section 4.3.2.1	Exploratory objective added to protocol planned analyses

Abbreviations: ALT = alanine aminotransferase; AST = aspartate aminotransferase; BL = baseline; eGFR = estimated glomerular filtration rate; ELF = enhanced liver fibrosis; FFA = free fatty acid; FIB-4 = Fibrosis-4; GGT = gamma-glutamyl transferase; HbA1c = glycated hemoglobin; HDL-C = high-density lipoprotein cholesterol; HOMA2-%B = homeostasis model assessment estimates steady state beta cell function; HOMA2-%S = homeostasis model assessment estimates steady state insulin sensitivity; hs-CRP = high-sensitivity C-reactive protein; IGT = impaired glucose tolerance; K-18 (M30) = keratin-18 M30 fragment; LDL-C = low-density lipoprotein cholesterol; MMRM = mixed model for repeated measures; NAFLD = non-alcoholic fatty liver disease; Pro-C3 = a fragment of the NH2-terminal propeptide of type III procollagen; T2DM = type 2 diabetes mellitus; TC = total cholesterol; TG = triglycerides; TNFα = tumor necrosis factor alpha; TZP = tirzepatide; UACR = urine albumin-to-creatinine ratio; VLDL = very low-density lipoproteins.

Unless otherwise stated, the assessment of secondary efficacy objective will be guided by the "efficacy" estimand, using the same population as for primary analysis.

4.5. Safety Analyses

Unless specified otherwise, safety assessments will be based on the SS (Table GPHZ.3.1). All data collected between randomization and the end date of study participation will be included, regardless of the adherence to study drug.

The statistical assessment of homogeneity of the distribution of categorical safety responses, between tirzepatide doses and placebo, will be conducted using Fisher's exact test, unless specified otherwise.

The mean change from baseline differences among treatments at all scheduled visits will be assessed via an MMRM using REML. Unless specified otherwise, if the safety parameter is assessed at the 72 weeks, then the model will include treatment group, visit and treatment-by-visit interaction and stratification factors as fixed effects, and baseline value of the safety parameter as a covariate. To model the covariance structure within participants, the unstructured covariance matrix will be used. If this model fails to converge, the covariance structures specified in Section 4.3.2.1 will be tested in order until met convergence. If the data does not warrant the MMRM model, then ANCOVA model will be conducted.

For selected safety parameters, time-to-first-event analysis via the Cox-proportional hazards model may be conducted. Participants without the event will be censored at the end of study participation. For participants experiencing the event, the "time-to-first-event" will be the time (in weeks) from first dose to first occurrence of the event.

Where necessary, the rate of events will be analyzed using a generalized linear mixed-effects model, assuming the number of events follow a negative binomial distribution and with treatment as a fixed effect. The logarithm of days during the active treatment period will be adjusted as an offset, to account for possible unequal treatment duration of follow-up among participants.

4.5.1. Extent of Exposure

Summary of duration of follow-up (defined as time in days from date of randomization to the date of the last study visit) and/or duration on study treatment (defined as time in days from date of first dose of study treatment to date of last dose of study treatment plus 7 days) will be provided by treatment group using data from SS.

For the summary of duration on study and study treatment, the frequency and percentage of participants falling into the following range will be summarized by planned treatment group as well:

- >0
- ≥4 weeks
- ≥8 weeks
- ≥12 weeks
- ≥16 weeks
- ≥20 weeks

- ≥24 weeks
- ≥36 weeks
- ≥48 weeks
- ≥52 weeks
- ≥72 weeks

In addition, the frequency and percentages of participants falling into the following exposure ranges for study and study treatment may be summarized by planned treatment group:

- 0 weeks
- >0 to <4 weeks
- ≥4 to <8 weeks
- ≥8 to <16 weeks
- ≥16 to <24 weeks
- ≥24 to <36 weeks
- ≥36 to <48 weeks
- ≥48 to <52 weeks
- ≥52 to <72 weeks
- ≥72 weeks

No p-values will be reported in these summaries as they are intended to describe the study populations, rather than test hypotheses about them.

4.5.2. Adverse Events

4.5.2.1. Treatment Emergent Adverse Events

A TEAE is defined as an event that first occurred, or worsened in severity, after baselines defined in Table GPHZ.4.1. The MedDRA LLT will be used in the treatment-emergent derivation. The maximum severity for each LLT during the baseline period, including ongoing medical history, will be used as baseline severity. For events with a missing severity during the baseline period, it will be treated as "mild" in severity for determining treatment-emergence. Events with a missing severity during the postbaseline period will be treated as "severe", and treatment-emergence will be determined by comparing to baseline severity.

For events occurring on the day of first taking study medication, the CRF-collected information (e.g., treatment emergent flag, start time of study treatment and event) will be used to determine whether the event was pre- versus post-treatment, if available. If the relevant information is not available, then the events will be counted as post-treatment.

The counts and percentages of participants with TEAEs will be summarized by treatment using MedDRA PT nested within SOC. Statistical comparisons will be applied at both the SOC and PT levels. Events will be ordered by decreasing frequency within SOC. SOC will be presented in alphabetical order. For events that are sex-specific, the denominator and computation of the percentage will include only participants from the given sex.

An overview of the number and percentage of participants who experienced a TEAE, TEAE related to study treatment, SAE, death, or discontinued from study and/or study treatment due to an AE will be summarized by treatment.

The counts and percentages of participants with TEAEs by maximum severity will be summarized by treatment using MedDRA PT. For each participant and TEAE, the maximum severity for the MedDRA PT is the maximum postbaseline severity observed from all associated LLTs mapping to the MedDRA PT. The maximum severity will be determined based on the non-missing severities. If all severities are missing for the defined postbaseline period of interest, it will show as missing in the table.

4.5.2.2. Common Adverse Events

The counts and percentages of participants with TEAEs, overall and common (common TEAEs occurred in ≥5% of participants in any treatment group before rounding), will be summarized by treatment using MedDRA PT. Events will be ordered by decreasing frequency.

4.5.3. Additional Safety Assessments

4.5.3.1. Deaths

A listing of all deaths during the study will be provided. The listing will include participant identification including the treatment, site number, date of death, age at the time of enrollment, sex, time from last dose of study drug to death (if participant had discontinued study drug), and primary cause of death.

4.5.3.2. Other Serious Adverse Events

The counts and percentages of participants who experienced an SAE (including deaths and SAEs temporally associated with or preceding deaths) during the postbaseline period will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC. The SOC will be in alphabetical order.

A listing of all SAEs will be provided. The listing will include treatment, participant identification including the site number, date of event, age at the time of enrollment, sex, MedDRA SOC and PT, AE start date, AE end date, severity, action taken related to study treatment, outcome, relationship to study drug, and the last dose date of study drug.

4.5.3.3. Other Significant Adverse Events

The counts and percentages of participants who discontinued from study drug or study due to an AE during the postbaseline period may be summarized by treatment group using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC.

4.5.3.4. Vital Signs

In the case that multiple measurements of an individual vital sign (e.g., sitting systolic blood pressure [BP]) are collected at the same visit, the mean of these measurements will be used for the analysis.

Vital signs (systolic BP, diastolic BP, and pulse) will be summarized by treatment group at each scheduled visit. Change from baseline to postbaseline values for vital signs will be summarized for participants who have both a baseline, and at least, 1 postbaseline result. Treatment differences in mean change from baseline for vital signs will be assessed using analysis model described in Section 4.3.2.1. Only scheduled measurements will be included in the mean change analyses.

Counts and percentages of participants with treatment-emergent abnormal (ie. high or low) vital signs (sitting systolic BP, diastolic BP, and pulse) at any time during the entire study (including the safety follow-up period) will be summarized by treatment for participants who have both baseline and at least 1 postbaseline result. A treatment-emergent abnormal high result is defined as a change from a value less than, or equal to, the high limit at baseline to a value greater than the high limit at any time that meets the specified change criteria during the postbaseline period. A treatment-emergent abnormal low result is defined as a change from a value greater than, or equal to, the low limit at baseline to a value less than the low limit at any time that meets the specified change criteria during the postbaseline period. To assess decreases, change from the minimum value during the baseline period to the minimum value during the postbaseline period will be used. To assess increases, change from the maximum value during the baseline period to the maximum value during the postbaseline period will be used. Both planned and unplanned measurements will be included in the analysis. The criteria for identifying participants with treatment-emergent vital sign abnormalities are listed in Table GPHZ.4.5.

Table GPHZ.4.5. Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurements

Parameter	Low	High
Systolic BP (mm Hg) (Supine or sitting – forearm at heart level)	≤90 and decrease from baseline ≥20	• ≥140 and increase from baseline ≥20; • ≥129 and increase from baseline ≥20.
Diastolic BP (mm Hg) (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.

4.5.3.5. Electrocardiograms

Summary statistics by treatment and by nominal visit will be provided for ECG parameters (heart rate, pulse rate [PR], QRS, QT, and QT corrected using Fridericia's correction factor [QTcF = QT / $RR^{0.333}$]). When the QRS is prolonged (e.g., a complete bundle branch block), QT and QTc should be used to assess ventricular repolarization. Thus, for a particular ECG, the following will be set to missing (for analysis purposes) when QRS is ≥120 msec: QT and QTcF.

Change from baseline to postbaseline values for ECG parameters will be summarized for participants who have both a baseline and at least 1 postbaseline result. Treatment differences in mean change from baseline for heart rate and PR will be assessed using the analysis model described in Section 4.3.2.1. Only planned measurements will be included in the mean change analyses.

The counts and percentages of participants who meet following criteria at any time during the entire study period (including the safety follow up time period) will be summarized by treatment group:

- treatment-emergent ECG abnormalities as listed in Table GPHZ.4.6.
- QT greater than 500 msec
- QTcF greater than 500 msec, and

Approved on 11 Jul 2023 GMT

- treatment-emergent increase from the maximum baseline in QTcF interval of greater than 30 msec, 60 msec, or 75 msec. Maximum baseline (see Table GPHZ.4.1 for details) will be the maximum non-missing observation in the baseline period. The maximum value during the treatment period will be analyzed. Scheduled and unscheduled measurements will be included.

Table GPHZ.4.6. Selected Categorical Limits for ECG Data

	Low		High	
Parameter	**Males**	**Females**	**Males**	**Females**
Heart Rate (bpm)	<50 and decrease ≥15	<50 and decrease ≥15	>100 and increase ≥15	>100 and increase ≥15
PR Interval (msec)	<120	<120	≥220	≥220
QRS Interval (msec)	<60	<60	≥120	≥120
QTcF (msec)	<330	<340	>450	>470

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; PR = pulse rate; QTcF = Fridericia's corrected QT interval.

4.5.3.6. Clinical Laboratory Evaluation

Descriptive summaries by treatment and by nominal visit, will be provided for the baseline and postbaseline values, as well as the change from baseline values. The associated descriptive will be presented in SI units and in CN units. Limits from the performing laboratory will be used to define low and high.

Observed and change from baseline values for each visit may be displayed in plots for participants who have both a baseline and at least 1 postbaseline planned measurement. Baseline will be the last non-missing observation prior to taking first study drug. Unplanned measurements will be excluded from plots.

A shift table will be provided including unscheduled measurements. The shift table will include the number and percentage of participants within each baseline category (low, normal, high, or missing) versus each postbaseline category (low, normal, high, or missing) by treatment. The proportion of participants shifted will be compared between treatments using Fisher's exact test.

A listing of abnormal findings will be created for laboratory analyte measurements, including qualitative measures. The listing will include participant identification, treatment group, laboratory collection date, study day, analyte name, and analyte finding.

The MMRM model will be used for the analysis during the treatment period including safety follow-up period for the continuous measurements for following lab tests:

- Calcitonin
- eGFR
- UACR
- Hepatic Enzymes (ALT, AST, ALP, TBL and Direct Bilirubin)
- Pancreatic Enzymes (p-amylase and lipase)

4.5.4. Patient Narratives

The following are "notable" events, from start of study drug through end of study participation (or data cutoff for the submission if earlier):

- Deaths
- SAEs
- Permanent discontinuations of study treatment due to AEs
- Pregnancy
- Severe MACE
- Severe dehydration
- Severe renal disorder
- Severe hypersensitivity

Patient narratives (participant level data and summary paragraph) will be provided for participants in the enrolled population with at least one notable event.

4.5.5. Special Safety Topics

For AESI or special safety topics, the counts and percentages of participants will be summarized by treatment and PT with decreasing frequency. Individual participant level data may be presented. Displays with individual participant level data may be created using various formats, such as a customized listing and/or a customized graphical patient profile. Adverse events of special interest are defined in each section of special safety topics, where applicable.

4.5.5.1. Exocrine Pancreas Safety

4.5.5.1.1. Pancreatic Enzyme

Observed pancreatic enzyme data (p-amylase and lipase) will be summarized by treatment. The counts and percentages of participants with maximum postbaseline pancreatic enzyme value exceeding the following thresholds will be provided by baseline pancreatic enzyme value (≤ ULN, > ULN), and postbaseline: ≤1 × ULN, (>1 to ≤3) × ULN, (>3 to ≤5) × ULN, (>5 to ≤10) × ULN, >10 × ULN.

Mean change for pancreatic enzymes may be evaluated using an MMRM model with log transformed (postbaseline measure/baseline measure), as response variable will be used to analyze each pancreatic enzyme.

4.5.5.1.2. Pancreatitis Events

Summaries of adjudicated and investigator-reported pancreatic events will be provided by treatment. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

Treatment emergent adjudication-confirmed pancreatitis will be considered as an AESI. Listing of participants with adjudicated pancreatitis may be provided if deemed necessary.

4.5.5.2. Gastrointestinal Adverse Events

4.5.5.2.1. Nausea, Vomiting, Diarrhea and Constipation

Summaries and analyses for incidence and severity of following events will be provided by each treatment group:

- Nausea
- Vomiting (including "vomiting" and "vomiting projectile")
- Diarrhea (including "diarrhea" and "diarrhoea")
- Constipation
- 3 events (nausea, vomiting and diarrhea) combined
- 4 events (nausea, vomiting, diarrhea and constipation) combined

Summary of the prevalence over time will also be presented. Time to the onset of nausea, vomiting, and diarrhea will be plotted.

4.5.5.2.2. Severe Gastrointestinal Events

Severe GI AEs (GI SOC) will be captured with AE-CRF form and serious cases will be captured with the SAE form. The PTs in the GI SOC MedDRA version at the time of database lock will be used to identify GI AEs, and only the PTs with serious/severe treatment-emergent cases will be considered as AESIs.

The counts and percentages of participants with severe/serious treatment-emergent GI events will be summarized by treatment.

4.5.5.3. Hepatobiliary Disorders

4.5.5.3.1. Hepatic Events

Severe/serious treatment-emergent hepatic events will be considered as AESI and summarized. The counts and percentages of participants with treatment-emergent potentially drug-related hepatic events will be summarized by treatment using the MedDRA PTs. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

4.5.5.3.2. Acute Gallbladder Disease

Events related to acute gallbladder disease will also be summarized by treatment groups by PT with decreasing frequency. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

4.5.5.3.3. Liver Enzymes

Common analyses for laboratory analyte measurements described in Section 4.5.3.6 are applicable for the liver enzyme related measurements. This section provides additional analyses for liver enzymes.

The counts and percentages of participants with the following elevations in hepatic laboratory tests at any time during the treatment period, and during the entire study including follow up period, will be summarized between treatment groups:

- The counts and percentages of participants with an ALT measurement ≥3 times (3 ×), 5 times (5 ×), and 10 times (10 ×) the central lab ULN will be summarized for all participants with a postbaseline value and for subsets based on various levels of baseline value.
    - participants whose non-missing maximum baseline value is ≤1 × ULN
    - participants whose maximum baseline is >1 × ULN, and
    - participants whose baseline values are missing.
- The counts and percentages of participants with an AST measurement ≥3 ×, 5 ×, and 10 × the central lab ULN will be summarized for all participants with a postbaseline value and for subsets based on various levels of baseline value, as described above for ALT.
- The counts and percentages of participants with a TBL measurement ≥2 × the central lab ULN during the treatment period will be summarized for all participants with a postbaseline value, and subset into 4 subsets:
    - participants whose non-missing maximum baseline value is ≤1 × ULN
    - participants whose maximum baseline is >1 × ULN, but <2 × ULN
    - participants whose maximum baseline value is ≥2 × ULN, and
    - participants whose baseline values are missing.
- The counts and percentages of participants with a serum ALP measurement ≥2 × the central lab ULN during the treatment period will be summarized for all participants with a postbaseline value and for the subsets based on various levels of baseline value, as described above for TBL.

Maximum baseline will be the maximum non-missing observation in the baseline period. The maximum postbaseline value will be the maximum non-missing value from the postbaseline period. Scheduled and unscheduled measurements will be included.

Plot of maximum total bilirubin versus maximum aspartate aminotransferase and maximum total bilirubin versus maximum alanine aminotransferase will be provided.

4.5.5.4. Hypoglycemia

The following categories in accordance with the 2020 American Diabetes Association position statement on glycemic targets (ADA 2020) will be defined in the database.

Glucose alert value (Level 1):

- **Documented symptomatic hypoglycemia** is defined as any time a participant feels that he/she is experiencing symptoms and/or signs associated with hypoglycemia, and has a BG level of <70 mg/dL (<3.9 mmol/L).
- **Documented asymptomatic hypoglycemia** is defined as any event not accompanied by typical symptoms of hypoglycemia, but with a measured BG <70 mg/dL (<3.9 mmol/L).

- **Documented unspecified hypoglycemia** is defined as any event with no information about symptoms of hypoglycemia available, but with a measured BG <70 mg/dL (<3.9 mmol/L).

Documented Clinically Significant Hypoglycemia (Level 2):

- **Documented symptomatic hypoglycemia** is defined as any time a participant feels that he/she is experiencing symptoms and/or signs associated with hypoglycemia and has a BG level of <54 mg/dL (<3.0 mmol/L).
- **Documented asymptomatic hypoglycemia** is defined as any event not accompanied by typical symptoms of hypoglycemia but with a measured BG <54 mg/dL (<3.0 mmol/L).
- **Documented unspecified hypoglycemia** is defined as any event with no information about symptoms of hypoglycemia available but with a measured BG <54 mg/dL (<3.0 mmol/L).

Severe Hypoglycemia (Level 3):

- **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. These episodes may be associated with sufficient neuroglycopenia to induce seizure or coma. PG measurements may not be available during such an event, but neurological recovery attributable to the restoration of PG to normal is considered sufficient evidence that the event was induced by a low PG concentration.

Other hypoglycemia categories:

- **Nocturnal hypoglycemia** is defined as any hypoglycemic event that occurs between bedtime and waking.

To avoid duplicate reporting, all consecutive hypoglycemic events occurring within a 1-hour period will be considered as a single hypoglycemic event.

Statistical summaries and analyses will exclude hypoglycemic events occurring after initiation of rescue antihyperglycemic medication. Both the incidence (percent of participants experiencing ≥1 episode) and the rate (episodes/patient/year) of level 2 or level 3 hypoglycemia, and level 3 hypoglycemia will be reported by treatment.

The incidence of hypoglycemic events will be analyzed using logistic regression with treatment and stratification factors as fixed effects. The rate of hypoglycemic episodes per patient year may be analyzed using a generalized linear mixed-effects model assuming the number of hypoglycemic episodes follows a negative binomial distribution with the mean modeled using stratification factors and treatment as fixed effects. A listing of hypoglycemic events will be provided.

If a hypoglycemic event meets the criteria of severe, the event would specifically be collected as an SAE. Serious hypoglycemia is defined by pharmacovigilance criteria and will also be captured with a SAE form.

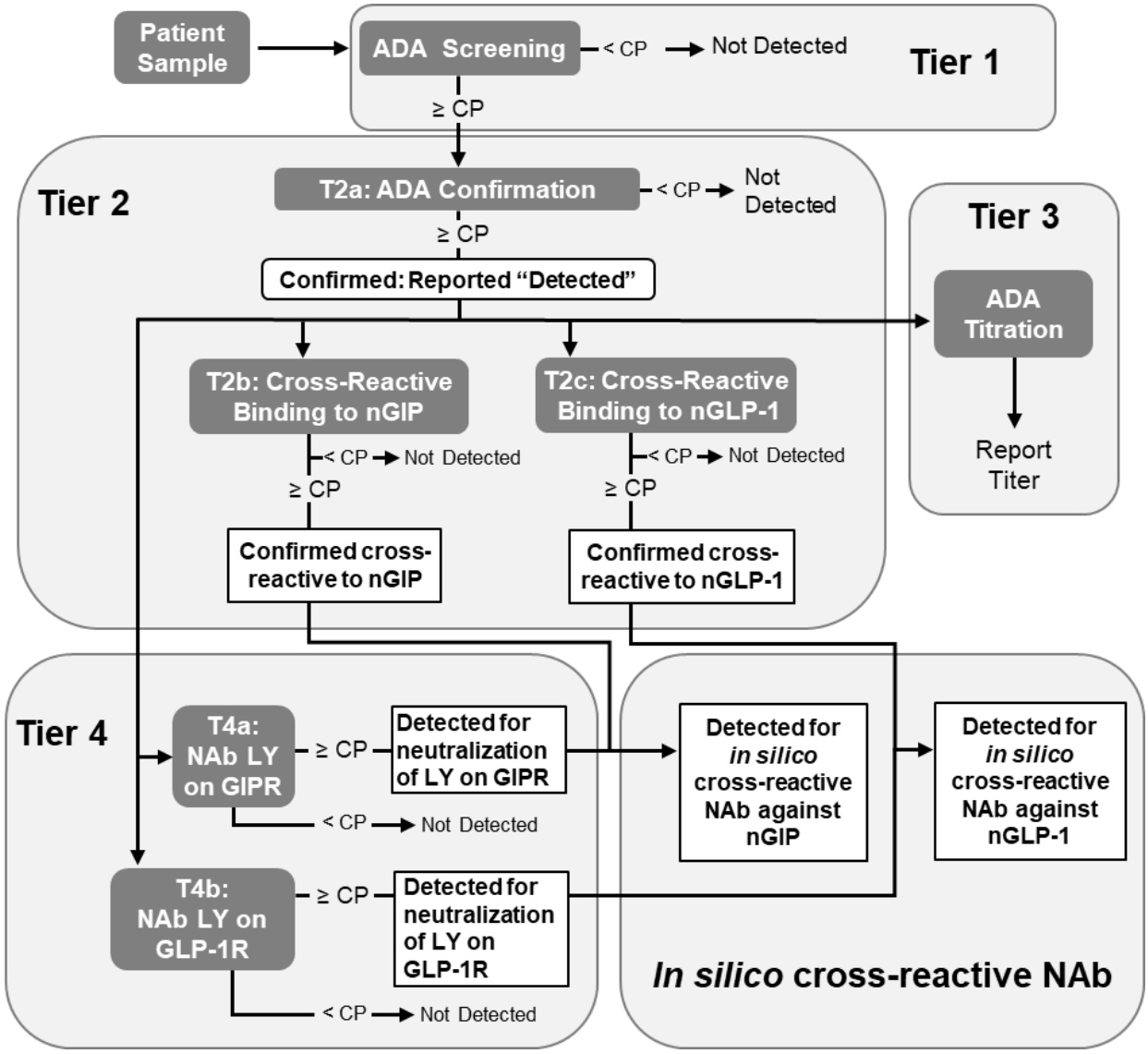
Severe/serious hypoglycemia will be considered as AESIs. The summaries of severe/serious hypoglycemia will be provided by treatment group, if deemed necessary.

4.5.5.5. Immunogenicity

4.5.5.5.1. Definitions of Sample ADA Status

At a high level, an individual sample is potentially examined multiple times, in a hierarchical procedure, to produce a sample ADA assay result, and potentially, multiple cross-reactive antibodies assay results, and multiple NAb assay results. The cut points used, the drug tolerance of each assay, and the possible values of titers are operating characteristics of the assay.

Figure GPHZ.4.1 details a flow chart that reflects the multi-tiered testing approach.



Abbreviations: ADA = anti-drug antibodies; CP = cut point; GIPR = the glucose-dependent insulinotropic polypeptide receptor; GLP-1R = glucagon-like peptide-1 receptor; LY = LY3298176 (tirzepatide); NAb = neutralizing antibodies; nGIP = native glucose-dependent insulinotropic polypeptide; nGLP-1 = native glucagon-like peptide-1.

Figure GPHZ.4.1. Flowchart of immunogenicity multitiered testing approach.

Table GPHZ.4.7 outlines results as reported from Tier 2a of the multi-tiered testing approach. Tier 4 results are reported similarly.

Table GPHZ.4.7. Sample Anti-Drug Antibodies (ADA) Assay Results

Sample Laboratory Result	Explanation
Detected	ADA are detected and confirmed.
Not Detected	The raw result as reported from the laboratory indicates not detected. The clinical interpretation of such results depends on other factors (see Table GPHZ.4.8).
NO TEST, QNS, etc	Sample exists but was unevaluable by the assay.

Abbreviations: ADA = anti-drug antibodies; QNS = quantity not sufficient.

It can be the case that the presence of high concentrations of tirzepatide will affect immunoassays, and conversely high levels of antibodies may affect the measurement of tirzepatide concentration. Thus, a tirzepatide drug concentration, assessed from a sample drawn at the same time as the ADA sample, plays a key role in clinical interpretation of a sample when the laboratory result is Not Detected (see Table GPHZ.4.8).

Table GPHZ.4.8. Sample Clinical ADA Interpretation Results

Sample Clinical Interpretation	Explanation
ADA Present	ADA assay result is Detected
ADA Not Present	ADA assay result is Not Detected and simultaneous drug concentration is at a level that has been demonstrated to not interfere in the ADA detection method (i.e., drug concentration is below the assay's drug tolerance level). For participants receiving placebo, drug concentration is not assessed and is assumed to be below the assay's drug tolerance level. If drug concentration was planned but is not available for a treatment-period sample, a Not Detected sample will be declared ADA Not present.
ADA Inconclusive	ADA assay result is Not Detected but drug concentration in the sample is ≥ the assay's drug tolerance level, which may cause interference in the ADA detection method.

Abbreviations: ADA = anti-drug antibodies.

All ADA present samples will be evaluated for cross-reactivity to native GIP (Tier 2b), cross-reactivity to native GLP-1 (Tier 2c), NAb LY (tirzepatide) on GIPR (Tier 4a), and NAb LY (tirzepatide) on GLP-1R (Tier 4b). If cross-reactivity ADA against native GIP is detected, the *in silico* assessment for cross-reactivity NAb against native GIP is evaluated, and if cross-reactivity ADA against GLP-1 is detected, the *in silico* assessment for cross-reactivity NAb against native GLP-1 is evaluated (Figure GPHZ.4.1).

Similar terminology to Table GPHZ.4.8 applies for each type of cross-reactive and NAb assay. Importantly, each of these is a distinct assay and, in general, has different assay operating characteristics.

The following are considered inconclusive for the NAb result:

- NAb LY on GIPR: if NAb result is not detected, and PK concentration is ≥ drug tolerance limit of the NAb LY on GIPR assay
- NAb LY on GLP-1R: if NAb result is not detected, and PK concentration is ≥ drug tolerance limit of the NAb LY on GLP-1R assay

An *in silico* method utilizing results from Tiers 2b and 2c, Tiers 4a and 4b, and tirzepatide concentrations is used to determine cross-reactive NAb against native GIP and GLP-1. The *in silico* method is outlined in the following table (Table GPHZ.4.9).

Table GPHZ.4.9. In Silico Classification for Cross-Reactive NAb

In Silico Classification	Cross-reactive Binding ADA Result	NAb Result	Circulating Tirzepatide Level (ng/mL)	*In Silico* Cross-reactive NAb Interpretation
Cross-reactive NAb to nGIP	Tier 2b: "Not Detected"	Tier 4a: "Not Detected" *or* "Detected" or N/A or Missing	Any Value or Missing	Not Present
	Tier 2b: "Detected"	Tier 4a: "Not Detected"	< drug tolerance limit of Tier 4a assay	Not Present
	Tier 2b: "Detected"	Tier 4a: "Not Detected"	≥ drug tolerance limit of Tier 4a assay	Inconclusive
	Tier 2b: "Detected"	Tier 4a: "Detected"	< drug tolerance limit of Tier 4a assay	Present
	Tier 2b: "Detected"	Tier 4a: "Detected"	≥ drug tolerance limit of Tier 4a assay	Present
Cross-reactive NAb to nGLP-1	Tier 2c: "Not Detected"	Tier 4b: "Not Detected" *or* "Detected" or NA or Missing	Any Value or Missing	Not Present
	Tier 2c: "Detected"	Tier 4b: "Not Detected"	< drug tolerance limit of Tier 4b assay	Not Present
	Tier 2c: "Detected"	Tier 4b: "Not Detected"	≥ drug tolerance limit of Tier 4b assay	Inconclusive
	Tier 2c: "Detected"	Tier 4b: "Detected"	< drug tolerance limit of Tier 4b assay	Present
	Tier 2c: "Detected"	Tier 4b: "Detected"	≥ drug tolerance limit of Tier 4b assay	Present

Abbreviations: ADA = antidrug antibodies; GIP = glucose-dependent insulinotropic polypeptide; GLP-1 = glucagon-like peptide-1; NAb = neutralizing antibodies; nGIP = native glucose-dependent insulinotropic polypeptide; nGLP-1 = native glucagon-like peptide-1; Tier 2b = cross-reactive ADA to nGIP; Tier 2c = cross-reactive ADA to nGLP-1; Tier 4a = NAb LY (tirzepatide) on GIP receptor; Tier 4b = NAb LY (tirzepatide) on GLP-1 receptor.

Note that, in the case of an ADA Inconclusive sample, each of the NAb and Cross-Reactive NAb assay results is taken to be Inconclusive.

Note also that any reference to an assay cut point and/or drug tolerance is population specific, and is subject to modification to study-specific parameters per regulatory guidance.

4.5.5.5.2. Definitions of Immunogenicity Assessment Periods

Immunogenicity Baseline Observations: Baseline period for immunogenicity assessment for each participant includes all observations up to baseline visit. In instances where multiple baseline observations are collected, to determine participant ADA status the last non-missing immunogenicity assessment up to first administration of study intervention is used to determine treatment-emergent status (see below).

Immunogenicity Postbaseline Period Observations: Postbaseline period observations for each participant include all observations after the first administration of study drug. There are 2 different periods listed below:

- The planned treatment period is defined as from the first dose of treatment to end of the treatment period.
- The entire postbaseline period is defined as from the first dose of treatment to the end of safety follow-up visit or date of study withdrawal.

4.5.5.5.3. Definitions of Participant ADA Status

Treatment-emergent (TE) ADA-evaluable participants: a participant with a nonmissing baseline ADA result and at least 1 nonmissing postbaseline ADA result.

Treatment-emergent ADA unevaluable participant: any participant who does not meet the evaluable criteria.

Baseline ADA Present (preexisting antibody): ADA detected in a sample collected up to the first dose date and time.

Baseline ADA Not Present: ADA is not detected, and the corresponding tirzepatide PK concentration is missing or below the drug tolerance limit in a sample collected up to the first dose date and time.

TE ADA positive (TE ADA+) participant: A participant who is evaluable for TE ADA is TE ADA+ if either of the following holds:

- The participant has baseline status of ADA Not Present and at least 1 postbaseline status of ADA Present with titer ≥2× MRD of the ADA assay.
- The participant has baseline and postbaseline status of ADA Present, with the postbaseline titer being 2 dilutions (4-fold) greater than the baseline titer. That is, the participant has baseline (B) status of ADA Present, with titer 1:B, and at least 1 postbaseline (P) status of ADA Present, with titer 1:P, with P/B ≥4.

As shown in Figure GPHZ.4.1, a titer is expected when ADA assay result is detected. On occasion, the corresponding assay cannot be performed, in which case a titer value will be imputed for the purpose of TE ADA determination. A baseline sample with detected ADA and

no titer is imputed to be the MRD (1:10) and a postbaseline sample with ADA detected and no titer is imputed to be one dilution above the MRD (1:20).

TE ADA-Inconclusive participant: a TE ADA-evaluable participant is TE ADA Inconclusive if ≥20% of the participant's postbaseline samples are ADA Inconclusive and all remaining postbaseline samples are ADA Not Present.

TE ADA-negative (TE ADA-) participant: a TE ADA-evaluable participant is TE ADA- when the participant is not TE ADA+ and not TE ADA Inconclusive.

For each NAb assay, the following is defined:

NAb-positive (NAb+) participant: a participant who is TE ADA+ and has a Nab+ sample in the postbaseline period.

NAb-Inconclusive participant: a participant who is TE ADA+, is not NAb+, and all samples that have TE ADA+ titer have a NAb-Inconclusive sample result.

NAb-negative (NAb-) participant: a participant is neither NAb+ or NAb inconclusive.

Unless specified otherwise, the above-mentioned definitions of NAb are applicable to all NAb analyses, including cross-reactive NAb analyses, and cross-reactive antibodies.

4.5.5.5.4. Analyses to be Performed

The count and proportion of participants who are TE ADA+ will be tabulated by treatment group, where proportions are relative to the number of TE ADA-evaluable participants, as defined above. The tabulation will include the count and proportion of participants with ADA Present at baseline and the count and proportion of TE ADA+ participants exhibiting each type of cross-reactive antibodies and NAb. This analysis will be performed for

- the planned treatment period, and
- the entire postbaseline period including safety follow-up.

The *in silico* classification for cross-reactive NAb will be summarized.

A summary will be provided of the count and percentage of tirzepatide-treated participants experiencing specific TEAE (see Table GPHZ.4.10) by TE ADA status (TE ADA+, TE ADA-, TE ADA Inconclusive) during the planned treatment period.

Table GPHZ.4.10. Adverse Events for Analysis with Immunogenicity Results

TEAE category	Criteria
Hypersensitivity reactions	Anaphylaxis SMQ (narrow or algorithm)
	Hypersensitivity SMQ (narrow)
	Angioedema SMQ (narrow)
	Severe Cutaneous Adverse Reaction SMQ (narrow)
	Vasculitis SMQ (narrow)
Injection site reactions	Injection site reaction HLT
	Infusion site reaction HLT
	Administration site reaction HLT

Abbreviations: HLT = high-level term; MedDRA = Medical Dictionary for Regulatory Activity; SMQ = standardized MedDRA query; TEAE = treatment-emergent adverse event.

A listing will be provided for all participants who had ADA Present at any time (including baseline) or had any specific TEAE (see Table GPHZ.4.10). This listing includes a time course of ADA (clinical interpretation result plus flags for samples meeting TE ADA+ criteria and for samples with cross-reactive antibodies and NAb present) along with the TEAE.

Cases of TE ADA that are associated with TEAEs of either severe/serious hypersensitivity or ISRs will be classified as AESIs.

Additional immunogenicity analyses as determined later may be presented. The relationship between antibody titers, the PK parameters, and PD response to tirzepatide may also be assessed.

4.5.5.6. Hypersensitivity Reactions

Two main analyses are performed in support of assessment of potential immediate hypersensitivity, including anaphylaxis as well as potential nonimmediate hypersensitivity.

Time Period A, of potential immediate hypersensitivity includes all TEAEs occurring from start of study drug administration up to 24 hours after end of study drug administration. For events without the hypersensitivity electronic case report form (eCRF), only date (no time) information is collected; if such event occurred on the same date as the study drug injection date, they will be included in Time Period A.

Time Period B, of potential non-immediate hypersensitivity, includes all TEAEs occurring more than 24 hours after the end of study drug administration, but prior to subsequent drug administration.

The counts and percentages of participants who experienced a hypersensitivity TEAE will be summarized by PT with decreasing frequency by treatment.

Analyses for both time periods are based on the following:

- Narrow and algorithm terms in Anaphylactic reaction SMQ (20000021)
- Narrow terms in Angioedema SMQ (20000024)
- Narrow terms in Severe cutaneous adverse reactions SMQ (20000020)
- Narrow terms in Hypersensitivity SMQ (20000214)
- Narrow terms in Vasculitis SMQ (20000174)

For the Anaphylactic reaction SMQ, each term is classified by scope (Narrow, Broad) and by category (A, B, C, D). All Narrow terms are category A, and all Broad terms are category B, C, or D. In addition to the usual Narrow and Broad searches, the SMQ defines an algorithm to further refine the cases of interest. For the analysis of the time period A, the Anaphylactic reaction SMQ algorithm will be included. The algorithm is based upon events that occur within Time Period A. The counts and percentages of participants who experienced a TEAE for the following will be analyzed for each of the 2 time periods:

- Any narrow term from any one of the 5 SMQs indicated above (i.e., combined search across narrow of all 5 SMQs)
- Any narrow term within each SMQ, separately (i.e., narrow SMQ search).

Within query, individual PTs that satisfied the queries will be summarized. Also, a single event may satisfy multiple SMQ, in which case the event contributes to every applicable SMQ.

4.5.5.6.1. Severe/Serious Hypersensitivity Reactions

The severe/serious cases of treatment-emergent hypersensitivity will be considered as AESIs. A summary with severe/serious hypersensitivity reactions may be provided, if deemed necessary.

4.5.5.7. Injection Site Reaction

Injection site reactions, incidence and rates, and related information reported via the "Injection Site Reactions" eCRF will be summarized by treatment. Information to be summarized includes the location of the reaction, timing of the reaction relative to study drug administration, and characteristics of the injection site reaction: erythema, induration, pain, pruritis, and edema.

Patient-based analysis and event-based analysis may be provided if necessary. The patient-based analysis summarizes all ISR questionnaire forms for an individual participant with a single statistic, typically an extreme value. This analysis allows each participant to contribute only once for each parameter, at the expense of a focus on the most extreme events. By contrast, the event-based analysis summarizes all ISR questionnaire forms received, without regard to individual participants. This provides characteristics of ISR events as a proportion of all events for which questionnaire responses were provided, at the expense of some potential bias due to differential contribution of individual participants to the analysis.

The counts and percentages of participants with treatment-emergent ISR will be summarized by treatment using the MedDRA PTs. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

The PT will be listed for summary in decreasing order of incidence for tirzepatide-treated participants.

4.5.5.7.1. Severe/Serious Injection Site Reactions

The severe/serious treatment-emergent injection site reactions based on TEAE searching criteria specified in Appendix 6 (Section 6.6) will be considered as AESI.

The counts and percentage of participants with severe/serious treatment-emergent ISRs will be summarized by treatment.

4.5.5.8. Major Adverse Cardiovascular Events

The following positively adjudicated MACE will be considered as AESIs:

- death due to cardiovascular AEs
- myocardial infarction,
- hospitalization for unstable angina,
- hospitalization for heart failure,
- coronary interventions (such as coronary artery bypass graft or percutaneous coronary intervention), and
- cerebrovascular events, including cerebrovascular accident (stroke) and transient ischemic attack.

The counts and percentages of participants with positively adjudicated MACE may be summarized by treatment.

In addition, MACE reported by investigator may also be summarized although a MACE reported by investigator and not positively adjudicated is not considered as AESI.

A listing of participants reporting MACE events, either reported by investigator or identified by the CEC will be provided.

4.5.5.9. Major Depressive Disorder/Suicidal Ideation or Behavior

The severe/serious treatment-emergent major depressive disorder/suicidal ideation or behavior will be captured as AESI. Adverse events will be searched using MedDRA PT terms. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

The counts and percentages of participants with TEAEs will be summarized by treatment group using MedDRA PT nested within SMQ. Events will be ordered by decreasing frequency nested within SMQ.

In addition to spontaneously reported AEs assessed by the investigator, suicidal ideation and behavior, and depression will be assessed through the use of the Columbia-Suicide Severity Rating Scale (C-SSRS) and the Patient Health Questionnaire (PHQ-9), respectively.

4.5.5.9.1. Patient Health Questionnaire (PHQ-9)

Total scores for the PHQ-9 range from 0 to 27 with total scores categorized as:

- None (not depressed): 0 – 4;
- Mild: 5 – 9;
- Moderate: 10 – 14;
- Moderately Severe: 15 – 19; and
- Severe: 20 – 27.

Shift tables will be provided showing the counts and percentages of participants within each baseline category (maximum value) versus each postbaseline category (maximum value) by treatment.

Additionally, the following 3 outcomes of interest will be compared between treatments (based on the maximum value during baseline and postbaseline):

- Any increase in depression category (i.e., worsening of depression); includes participants in the none, mild, moderate, or moderately severe category during baseline and with at least 1 postbaseline measurement;
- Increase from No or Mild Depression to Moderate, Moderately Severe or Severe Depression; includes participants in the none or mild depression category during baseline and with at least 1 postbaseline measurement; and
- Increase from Mild or Moderate Depression to Moderately Severe or Severe Depression; includes participants in the mild or moderate depression category during baseline and with at least 1 postbaseline measurement.

4.5.5.9.2. Suicidal Ideation and Behavior Solicited Through C-SSRS

Suicide-related thoughts and behaviors occurring during the entire study period, based on the C-SSRS, will be summarized by treatment group. In particular, for each of the following suicide-related events, the counts and percentages of participants with the event will be summarized by treatment group:

- Died by suicide,
- Nonfatal suicide attempt,
- Interrupted attempt,
- Aborted attempt,
- Preparatory acts or behavior,
- Active suicidal ideation with specific plan and intent,
- Active suicidal ideation with some intent to act without specific plan,
- Active suicidal ideation with any methods (no plan) without intent to act,
- Nonspecific active suicidal thoughts,
- Wish to be dead, and
- Non-suicidal, self-injurious behavior.

In addition, the counts and percentages of participants who experienced at least 1 of the composite measures will be presented. The participants with at least 1 postbaseline C-SSRS assessment are included. The composite measure is determined at each assessment by the "yes" or "no" responses in the following C-SSRS categories by the study participant:

- Category 1 – Wish to be Dead;
- Category 2 – Non-specific Active Suicidal Thoughts;
- Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act;
- Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan;
- Category 5 – Active Suicidal Ideation with Specific Plan and Intent;
- Category 6 – Preparatory Acts or Behavior;
- Category 7 – Aborted Attempt;
- Category 8 – Interrupted Attempt;
- Category 9 – Actual Attempt (non-fatal);
- Category 10 – Completed Suicide.

Composite endpoints of suicidal ideation and suicidal behavior based on the above categories are defined below:

- **Suicidal ideation**: A "yes" answer at any time during study to any 1 of the 5 suicidal ideation questions (Categories 1 – 5) on the C-SSRS;
- **Suicidal behavior**: A "yes" answer at any time during study to any 1 of the 5 suicidal behavior questions (Categories 6 – 10) on the C-SSRS;
- **Suicidal ideation or behavior**: A "yes" answer at any time during study to any 1 of the 10 suicidal ideation and behavior questions (Categories 1 – 10) on the C-SSRS.

A listing containing data for each participant with suicidal ideation, suicidal behavior, or non-suicidal self-injurious behavior during the study may be provided. Data from all visits are displayed, regardless of a "yes" or "no" answer, for participants with any "yes" answer for suicidal ideation, suicidal behavior, or self-injurious behavior without suicidal intent.

4.5.5.10. Malignancy

Treatment-emergent malignancy will be considered as an AESI. The counts and percentages of participants with treatment emergent malignancy will be summarized by treatment and PT and ordered by decreasing frequency. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

4.5.5.11. Renal Safety

Laboratory measures related to renal safety will be analyzed as specified for laboratory measurements in Section 4.5.3.6.

Two shift tables examining changes in renal function from baseline to postbaseline will be created. A min-to-min shift table of eGFR estimated by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation with units mL/min/1.73m^2, using categories (<30, ≥30 to <45, ≥45 to <60, ≥60, and ≥90 mL/min/1.73m^2). A max-to-max shift table of urine albumin-to-creatinine ratio (UACR), using the categories UACR <30 mg/g, 30 mg/g ≤ UACR ≤300 mg/g, UACR >300 mg/g (respectively, these represent normal, microalbuminuria, and macroalbuminuria).

Mixed model repeated measure analyses for eGFR and UACR will be provided. Log transformation will be performed for UACR.

4.5.5.11.1. Acute Renal Events

Acute renal events associated with chronic renal failure exacerbation will be captured.

Severe/serious renal events from the SMQ search below will be considered AESIs.

The counts and percentages of participants with acute renal events will be summarized by treatment by using the MedDRA PTs contained in any of the following SMQs:

- **Acute renal failure:**
  - Narrow terms in Acute renal failure SMQ (20000003)
- **Chronic kidney disease:**
  - Narrow terms in Chronic kidney disease SMQ (20000213)

In addition, a listing of participants with treatment-emergent acute renal events may be provided, if deemed necessary.

4.5.5.11.2. Dehydration

Dehydration events will be captured in the following SMQ and summarized. Severe/serious dehydration events will be considered as AESI:

- Narrow terms in Dehydration SMQ (20000232)

The counts and percentages of participants with dehydration will be summarized by treatment and PT and ordered by decreasing frequency. A listing of participants with treatment-emergent dehydration events may be provided, if deemed necessary.

4.5.5.12. Thyroid Safety Monitoring

4.5.5.12.1. Calcitonin

Observed calcitonin data (a thyroid-specific laboratory assessment) will be summarized by treatment and nominal visit.

The counts and percentages of participants with a maximum postbaseline calcitonin value in the following thresholds will be provided by treatment and baseline calcitonin value: (≤20 ng/L, >20 ng/L to ≤35 ng/L, >35 ng/L). Postbaseline : ≤20 ng/L, >20 ng/L to ≤35 ng/L, >35 ng/L to ≤50 ng/L, >50 ng/L to ≤100 ng/L, and >100 ng/L.

4.5.5.12.2. C-Cell Hyperplasia and Thyroid Malignancies

Treatment-emergent thyroid malignancies and C-cell hyperplasia will be considered as AESI. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

The counts and percentages of participants with treatment-emergent thyroid C-cell hyperplasia and malignancies will be summarized by treatment and PT ordered with decreasing frequency. In addition, a listing of participants with treatment-emergent thyroid C-cell hyperplasia and neoplasms may be provided if deemed necessary.

4.5.5.13. Treatment-Emergent Arrhythmias and Cardiac Conduction Disorders

Severe/serious treatment-emergent arrhythmias and cardiac conduction disorders will be considered as AESI.

The treatment-emergent arrhythmias and cardiac conduction disorders events will be identified using the MedDRA PTs. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

The counts and percentages of participants with treatment emergent arrhythmias and cardiac conduction disorders will be summarized by treatment and PT nested within SMQ. The PT will be ordered with decreasing frequency within SMQ. A listing of participants with treatment-emergent arrhythmias and cardiac conduction disorders may be provided if deemed necessary.

4.5.5.14. Treatment of Overdose

A listing of patients reporting AEs related to overdosing of tirzepatide may be provided.

4.5.5.15. Abuse Potential

To identify AE terms suggestive of abuse liability potential, narrow terms from the SMQ Drug abuse and dependence (20000101) will be used. The counts and percentages of participants will be summarized by treatment group with decreasing frequency.

4.6. Other Analyses

4.6.1. Health Outcomes

The patient-reported outcome questionnaires will be analyzed using the mITT population on the EAS, unless specified otherwise.

Item-level missingness is dealt with as per the instrument developers' instruction.

Additional psychometric analyses will be performed by the Value Evidence Outcomes group at Lilly and documented in a separate analysis plan.

4.6.1.1. Short-Form-36 Health Survey Version 2, Acute Form

Per copyright owner, the QualityMetric Health Outcomes™ Scoring Software (PRO_CoRe V2.0) will be used to derive the following domain and component scores:

- Mental Component Summary (MCS)
- Physical Component Summary (PCS)
- Physical Functioning domain (PF)
- Role-Physical domain (RP)
- Bodily Pain domain (BP)
- General Health domain (GH)
- Vitality domain (VT)
- Social Functioning domain (SF)
- Role-Emotional domain (RE), and
- Mental Health domain (MH).

For each above domain and component summary scores parameter, the raw scores will be transformed into the domain scores (t-scores) and the following analyses for the actual value and change from baseline value will be conducted:

- descriptive summaries by treatment group, and
- analysis described in Table GPHZ.4.11.

If data allowed, analysis for SF-36 physical function domain score analysis described in Table GPHZ.4.11 will be conducted to evaluate the treatment effect in participants who have limitations in physical function at baseline, which is defined as PGIS response at baseline of "moderately limited," "very much limited," or "extremely limited."

4.6.1.2. Impact of Weight on Quality of Life-Lite Clinical Trials

The following parameters will be included from IWQOL-Lite-CT:

- IWQOL Lite CT total score (all items: items 1 through 20)

- Physical Function composite score (5 items: items 1 through 3, 16, 17)
- Physical composite score (7 items: item 1 through 5, 16, 17), and
- Psychosocial composite score (13 items: item 6 through 15, 18, 19, 20).

IWQOL-Lite-CT total and composite scores range from 0 to 100, with higher scores reflecting better levels of functioning.

IWQOL-Lite-CT scores are computed according to the IWQOL-Lite scoring rules (Kolotkin et al. 2002) as following:

- Each composite raw score will be calculated if a minimum of 50% of the items for that composite has a non-missing value; the total score will be calculated if a minimum of 75% of all 20 items has a non-missing value.
    - physical composite score: 4 of 7 items
    - physical function composite score: 3 of 5 items
    - psychosocial composite score: 7 of 13 items
    - IWQOL-Lite-CT total score: 15 of 20 items
- If the minimum required number of items is answered then:
    - The average of the valid non-missing responses corresponding to the items in the total or each composite will be calculated (1 = "never" or "not at all true" and 5 = "always" or "completely true").
    - The composite score will be then calculated by transforming the raw composite score to the 0 (worst)-to-100 (best) metric using the following formula for every patient at each time point:

$$100\ (S_{max} - C_{avg})/(S_{max} - S_{min})$$

        - C_{avg} is the raw average score of all non-missing item responses in the composite; this average must be a number between 1 and 5, inclusive
        - S_{max} is the maximum possible raw score value (i.e., 5)
        - S_{min} is the minimum possible raw score value (i.e., 1)
        - Inserting the maximum and minimum possible score values, the formula is reduced to $100\ (5 - C_{avg})/4$.

For total and each composite score, the following analyses for the actual value and change from baseline value will be conducted:

- descriptive summaries by treatment group and
- ANCOVA analysis described in Table GPHZ.4.11.

4.6.1.3. EQ-5D-5L

Generic health-related quality of life will be assessed using the EuroQol – 5 Dimension – 5 Level (EQ-5D-5L) (EuroQoL Group 2015). The EQ-5D-5L is a standardized 5-item instrument for use as a measure of health outcome. It provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care as well as population health surveys. The EQ-5D-5L comprises 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The 5L version, introduced in 2005, scores each dimension at 5 levels (no problems, slight problems, moderate problems, severe problems, unable to perform/extreme problems), for a total of 3125 possible health states. In addition to the health profile, a single health state index value can be derived based on a formula that attaches weights to each of the levels in each dimension. This index value ranges between <0 (where 0 is a health state equivalent to death; negative values are valued as worse than dead) to 1 (perfect health) (Dolan 1997). In addition, the EQ visual analog scale (VAS) records the respondent's self-rated health status on a vertical graduated (0 to 100) scale. In conjunction with the health state data, it provides a composite picture of the respondent's health status. The EQ-5D-5L is used worldwide and is available in more than 170 different languages. Details on the instrument, and scoring, organizing, and presenting the data collected will follow C2H guidelines (Ikeda et al. 2015; C2H 2019).

4.6.1.4. Patient Global Impression of Status for Physical Activity

Study participants will be asked to complete a PGIS item. This is a patient-rated assessment of current health limitation and is rated on a 5-point scale ranging from "1- not at all limited" to "5- extremely limited."

The counts and percentages of participants for Patient Global Impression of Status for Physical Activity (PGIS) response categories at each time point will be summarized by nominal visit and by treatment. A shift table from baseline to postbaseline of 5 PGIS response categories at each postbaseline visit by treatment will be created.

4.6.1.5. Health Outcome Analysis

Table GPHZ.4.11. Health Outcome Measures

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Change in patient-reported physical functioning (72 weeks)	Mean change from BL in SF 36v2 acute form physical functioning domain score	Efficacy analysis set	ANCOVA model in Section 4.4.2.1	
	Mean change from BL in IWQOL-Lite-CT Physical Function score	Efficacy analysis set	ANCOVA model in Section 4.4.2.1	
Change in patient-reported health status (72 weeks)	Mean change from BL in EQ-5D-5L score • Index score • VAS score	Efficacy analysis set	ANCOVA model in Section 4.4.2.1	
Change in SF-36v2 acute scores other than physical functioning (72 weeks)	Mean change from BL in • role-physical score • bodily pain score • general health score • vitality score • social functioning score • role-emotional score • mental health score • physical component summary • mental component summary	Efficacy analysis set	ANCOVA model in Section 4.4.2.1	Exploratory objective
Change in IWQOL-Lite-CT scores other than physical functioning (72 weeks)	Mean change from BL in • physical score • psychosocial score • total score	Efficacy analysis set	ANCOVA model in Section 4.4.2.1	Exploratory objective
Change in Patient Global Impression of Status (72 weeks)	Shift of the ordinal response category from BL	Efficacy analysis set	Shift table described in Section 4.6.1.4	Exploratory objective

Abbreviations: ANCOVA = analysis of covariance; BL = baseline; EQ-5D-5L = EuroQol - 5 Dimension - 5 Level; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite-Clinical Trials Version; SF-36v2 = Short Form 36-item Health Survey Version 2; VAS = Visual Analog Scale.

4.6.2. Dietary Evaluation

All patients who additionally agree to participate in taking dietary evaluation will be in the analysis population for dietary evaluation (meal intake and appetite sensation) analyses. Dietary evaluation analyses will be guided by the efficacy estimand in EAS.

4.6.2.1. Meal Intake

Meal record will be collected before each meal (including breakfast, lunch, dinner, and snacks) for a day. In case if patients cannot finish them all, meal records after the meal will be collected. Meal records will be collected on 3 consecutive or non-consecutive days during the week (at least 1 day should be working day and at least 1 day should be resting day) immediately before the visits: Visit 3 (Week 0), Visit 9 (Week 24), Visit 16 (Week 52), and Visit 21 (Week 72). The Lilly-designated vendor will calculate the volume of total fats (g), total carbohydrates (g), total protein (g), and total caloric intake (kcal) based on the meal record.

For each participant and scheduled visit, daily parameters (sum of all days for each parameter at breakfast, lunch, dinner, and snack divided by the number of days) will be derived for following:

- Total caloric intake per day (kcal)
- Total fats per day (g)
- Total carbohydrates per day (g)
- Total protein per day (g)

For each participant and scheduled visit, daily caloric intake will be derived for following:

- Total fats per day (kcal) = 9 (kcal/g) × total fats per day (g)
- Total carbohydrates per day (kcal) = 4 (kcal/g) × total carbohydrates per day (g)
- Total protein per day (kcal) = 4 (kcal/g) × total protein per day (g)

For each participant and scheduled visit, percent of each component will be derived for following:

- Percent of fats per day (%) = 100 × total fats per day (kcal) / total caloric intake per day (kcal)
- Percent of carbohydrates per day (%) = 100 × total carbohydrates per day (kcal) / total caloric intake per day
- Percent of protein per day (%) = 100 × total protein per day (kcal) / total caloric intake per day

For each participant and scheduled visit, snack caloric intake per day will be derived by:

- Sum of all days for caloric intake at snack / number of days

Analysis method is described in the Section 4.6.2.3.

4.6.2.2. Appetite Sensation

The subjective rating of appetite sensations is measured by a 100-mm VAS for parameters of hunger, fullness, satiety, and prospective food consumption including snacks. The meal record will be collected on 3 consecutive or non-consecutive days during the week (at least 1 day should be working day and at least 1 day should be resting day) immediately before the visits: Visit 3 (Week 0), Visit 9 (Week 24), Visit 16 (Week 52), and Visit 21 (Week 72). For each meal, including breakfast, lunch, dinner and snacks, participant should complete the VAS 10 minutes before the meal and immediately after the meal. The timing of the record should be on the same date as the meal intake record. In case of frequent snacking on each recording day, if a participant consumes several snacks a day, they will be asked to complete the meal intake record for all snacks, but will only be required to complete VAS measurement for the first afternoon snack. The VAS is a validated tool to assess appetite sensation parameters (Flint et al. 2000). The VAS is presented as a 10-cm (100-mm) line, anchored by verbal descriptors, usually "extremely" and "not at all." Participants are required to rate their subjective sensations on four 100-mm scales combined with the questions:

- "How hungry do you feel right now?"
- "How satisfied do you feel right now?"
- "How full do you feel right now?"
- "How much food do you think you could eat right now?"

A staff member will use a caliper to measure the distance from 0 to the mark that the participant placed on the VAS and record the measurement in the source document.

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

{[Satiety] + [Fullness] + (100 – [Prospective food consumption]) + (100 – [Hunger])} / 4.

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

Analysis method is described in the Section 4.6.2.3.

4.6.2.3. Dietary Evaluation Analysis

Table GPHZ.4.12. Dietary Evaluation Measures

Analysis variable	Criteria/Measures	Analysis set	Analysis method	Additional Information
Change in dietary evaluation based on the meal record (24, 52, and 72 weeks)	Mean change from BL in • Total fats (g) • Total carbohydrates (g) • Total proteins (g) • Total caloric intake (kcal) • Total fats (kcal) • Total carbohydrates (kcal) • Total proteins (kcal) • Total fats (%) • Total carbohydrates (%) • Total proteins (%) • Snack caloric intake (kcal)	Efficacy analysis set – agreed dietary evaluation	MMRM model in Section 4.3.2.1	
Change in dietary evaluation based on appetite sensation (24, 52, and 72 weeks)	Mean change from BL in • Satiety • Fullness • Prospective food consumption • Hunger • Overall appetite score	Efficacy analysis set – agreed dietary evaluation	MMRM model in Section 4.3.2.1	

Abbreviations: BL = baseline; MMRM = mixed model for repeated measures.

4.6.3. Subgroup analyses

4.6.3.1. Subgroup Analysis of Efficacy Endpoints

Efficacy subgroup analyses will be guided by the efficacy estimand in EAS. Subgroup analyses by the following baseline characteristics will be provided:

- age (<65, ≥65 years)
- sex (female and male)
- baseline BMI (<35, ≥35 kg/m^2)
- baseline BMI (≥27 and <30, ≥30 and <35, ≥35 and <40, ≥40 kg/ m^2)
- glycemic status at screening (normoglycemia versus IGT)
- hyperlipidemia status at screening (TG; ≥150 and <150 mg/dL)
- NAFLD status at screening (HFF; ≥5% and <5%), analysis population will be the mITT excluding subjects with missing HFF measure at screening

For the reporting in the CSR, the outcome measures for the subgroup analyses will include:

- Percent change in body weight from randomization at 72 weeks.
- Proportion of participants achieving at least 5% body weight reduction from randomization to 72 weeks.

For the other purpose (eg. the publication), the outcome measures for the subgroup analyses will include:

- Proportion of participants achieving at least 10%, 15% and 20% body weight reduction from randomization to 72 weeks.
- Change in waist circumference (cm) from randomization to 72 weeks.
- Change in VAT area (cm^2), SAT area (cm^2) and VAT/SAT ratio from randomization to 72 weeks.
- Change in HFF (%) from randomization to 72 weeks.
- Change in Short-Form-36 Health Survey version 2 from randomization to 72 weeks.
- Change in IWQOL-Lite-CT Physical Function score from randomization to 72 weeks.

For the percentage change in body weight from randomization at 72 weeks, for each subgroup analyses aforementioned, the following analyses will be conducted:

- Conduct MMRM model on the subgroup only with terms of treatment group, visit, treatment-by-visit-interaction and stratification factors as fixed effects, and baseline body weight as a covariate. Variance-covariance structure for within-patient errors will be same as Section 4.3.2.1.

- Full MMRM model: treatment group, visit, subgroup, treatment-by-visit-interaction, treatment-by-subgroup-interaction, subgroup-by-visit-interaction, treatment-visit-subgroup-interaction, and stratification factors as fixed effects, and baseline body weight as a covariate. Variance-covariance structure for within-patient errors will be same as Section 4.3.2.1.

For the percentage of participants achieving at least 5% body weight reduction at 72 weeks, for each subgroup analyses aforementioned, the following analyses will be conducted:

- Conduct logistic regression model on the subgroup only with terms of treatment group, and stratification factors as fixed effects, and baseline body weight as a covariate. Variance-covariance structure for within-patient errors will be same as Section 4.3.2.2.
- Full logistic regression model: treatment group, subgroup, treatment-by-subgroup-interaction, and stratification factors as fixed effects, and baseline body weight as a covariate. Variance-covariance structure for within-patient errors will be same as Section 4.3.2.2.

4.6.3.2. Subgroup Analysis of Safety Endpoints

Safety subgroup analyses will be guided by the SS group. Subgroup analyses by the following baseline characteristics will be provided:

- age (<65, ≥65 years)
- sex (female and male)
- baseline BMI (≥27 and <30, ≥30 and <35, ≥35 and <40, ≥40 kg/ m^2)
- baseline eGFR (<60, ≥60 mL/min/1.73m^2)

The outcome measures for the subgroup analyses will include:

- TEAE

4.7. Other Exploratory Analyses

Exploratory analyses in this section will not be used for the CSR/CTD, but for the other purpose (eg. the publication).

To describe subject status regarding 3 defined ORHPs (IGT, hyperlipidemia and NAFLD) from baseline to Week 72, a Sankey diagram will be generated.

Scatter plot of creatinine and change in BMI (or body weight (kg)) will be created.

To explore relationships between change in PRO and the percentage of participants achieving at least 5% (and 7%, 10%, 15% or 20%) body weight reduction at 72 weeks, cross tabulation may be generated. PRO items include:

- SF36v2, acute form
    - Mental Component Summary (MCS)
    - Physical Component Summary (PCS)
    - Physical Functioning domain (PF)

- IWQOL-Lite-CT
    - Total score
    - Physical function composite score
    - physical score
    - Psychosocial composite score

The Analysis for achievement of ALT ≤30 U/L using longitudinal logistic regression model in Section 4.4.2.3 will be provided.

The analysis of the change in NFS (NAFLD fibrosis score) will be conducted.

- NFS = -1.675 + 0.037×age + 0.094×BMI (kg/m^2) + 1.13×IFG/diabetes (yes=1, no=0) + 0.99×AST/ALT - 0.013×blood platelet (×10^9/L) - 0.66×albumin (g/dL). (Tokushige et al. (2020))

4.8. Interim Analyses

Not applicable.

4.9. Changes to Protocol-Planned Analyses

Efficacy analysis set for imaging data has been modified to accommodate the allowance of measurement schedule.

Added following exploratory objectives:

- Improvement of hyperuricemia (defined by uric acid ≤7.0 mg/dL).
- Improvement of hypertension: Diastolic blood pressure <80 mmHg or Systolic blood pressure <130 mmHg.
- Achievement of ALT ≤30 U/L.
- Changes in NAFLD fibrosis score.

5. Sample Size Determination

Approximately, 348 participants will be screened to achieve 261 randomly assigned to study intervention (approximately 87 participants per intervention group).

The sample size determination assumes that evaluation of superiority of tirzepatide 10 mg and tirzepatide 15 mg to placebo will be conducted in parallel, each at a 2-sided significance level of 0.025 using a 2-sample t-test. Additionally, a difference of at least 11% mean body weight percentage reduction from randomization at 72 weeks for tirzepatide 10 mg and/or tirzepatide 15 mg compared with placebo, a common SD of 10%, and a dropout rate of 25% are assumed for statistical power calculations. Under the assumptions above, randomizing 261 participants in a 1:1:1 ratio to tirzepatide 10 mg, tirzepatide 15 mg, and placebo provides more than 90% power to demonstrate superiority of each tirzepatide dose to placebo.

The chosen sample size and randomization ratio also provides >90% power to establish superiority of tirzepatide 10-mg and/or tirzepatide 15-mg doses to placebo in terms of proportion of participants achieving at least 5% body weight reduction at 72 weeks, conducted in parallel using a Fisher's exact test, each at a 2-sided significance level of 0.025, assuming 25% of placebo-treated participants and 90% of tirzepatide-treated participants achieve the goal and a dropout rate of 25%.

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics of Study Population

6.1.1. Patient Characteristics

A listing of participant demographics for all randomized participants will be provided. All demographic and baseline clinical characteristics will be summarized by study treatment for all randomized participants. Baseline demographic and clinical characteristics of special interest include but are not limited to: age (years), sex (female, male), employment status, education level, marital status, household size, height (cm), weight (kg), BMI (kg/m^2), waist circumference (cm), age group (<65 years, ≥65 years), BMI group (<35, ≥35 kg/m^2), BMI group (<30, ≥30 and <35, ≥35 and <40, ≥40 kg/m^2), blood pressure (mmHg), pulse rate (bpm), eGFR (mL/minutes/1.73 m^2), eGFR group (<60, ≥60 mL/minutes/1.73 m^2) , eGFR group (≥30 and <45, ≥45 and <60, ≥60 and <90, ≥90 mL/minutes/1.73 m^2), triglyceride (mg/dL), triglyceride group (<150, ≥150 mg/dL), HFF (%), HFF group (<5, ≥5%), uric acid (mg/dL), uric acid group (<7, ≥7 mg/dL), glycemic status (IGT or normoglycemia), status of 3 defined ORHPs (IGT, hyperlipidemia and NAFLD) at screening, the number of 3 defined ORHPs (1, 2 or 3).

6.1.2. Historical Illnesses and Preexisting Conditions

The number and percentage of participants with historical illnesses and preexisting conditions will be summarized by treatment group using MedDRA PTs nested within SOC. System organ class will be in alphabetical order. Conditions (i.e., PTs) will be ordered by decreasing frequency within SOC. A preexisting condition is defined as the condition/event recorded on the *Preexisting Conditions and Medical History* eCRF page or on the *Prespecified Medical History* eCRF page with a start date prior to the date of informed consent, and no end date (i.e., the event is ongoing) or an end date on or after the date of informed consent. A historic illness is defined as an event with start date and stop date that are prior to the informed consent. The AEs occurring prior to first dose of study treatment will be included in the preexisting condition for reporting purpose.

6.1.3. Obesity-Related Health Problems

The numbers and percentages of participants with obesity-related health problems defined by the JASSO guideline, on the special *Prespecified Medical History: 11 Obesity-Related Health Problems* eCRF page, will be summarized by study treatment for all randomized participants and each BMI group (<30, ≥30 and <35, ≥35 kg/m^2).

6.1.4. Concomitant Medications

Summaries of Preferred Names of concomitant medications at baseline with number and percentage of patients sorted by decreasing frequency will be generated by treatment group for SS group. Summary of concomitant medication from randomization through safety follow-up will be generated as well.

Additionally, concomitant medications of interest (as defined below) will be summarized. The concomitant therapies will be mapped using the World Health Organization (WHO) DRUG dictionary in the clinical trial database and will be further classified using Anatomic-Therapeutic-Chemical (ATC) codes for reporting purposes. Concomitant therapy will be categorized by the ATC Level 4.

Concomitant medications of interest include the following:

- baseline use of
    - antihypertensive therapy, by type/class
    - lipid lowering therapy, by type/class
- utilization after randomization through safety follow-up:
    - medicines that cause weight gain
    - antihyperglycemic medication for the treatment of diabetes for participants who develop T2DM during the study (antihyperglycemic medication for the treatment of prediabetes is not allowed pre protocol)
    - antidiarrheal medication
    - antiemetic medication
    - anticonstipation medication
- status in changes to baseline medication in post-randomization (in term of type/class and dose):
    - antihypertensive therapy
    - lipid lowering therapy
    - antihyperuricemic therapy

Status in change of concomitant therapy will be categorized as:

- No use at both baseline and postbaseline period
- Increased (including increasing dose, starting of therapy)
- Not changed
- Decreased (including decreasing dose, stopping of therapy)
- Cannot be determined

6.2. Appendix 2: Treatment Compliance

If data warrant, the counts and percentages of participants who follow the planned escalation scheme, have temporary drug interruption, or have dose de-escalation/re-escalation will be summarized for tirzepatide treatment group. In addition, the proportion of participants receiving 2.5 mg, 5 mg, 7.5 mg, 10 mg, 12.5 mg, or 15 mg may be presented by visit during the dose escalation period.

Treatment compliance will be defined as taking at least 75% of the scheduled tirzepatide doses. Compliance over the treatment period will be calculated using the number of doses administered

(regardless of the actual dose in mg administered) divided by the total number of doses expected to be administered ×100 over the corresponding study period, respectively. Treatment compliance will be summarized descriptively over the treatment period by treatment using the mITT population.

6.3. Appendix 3: Important Protocol Deviations

Important protocol deviations are identified in the Trial Issues Management Plan (TIMP). A listing and a summary of important protocol deviations by treatment will be provided at the end of study.

6.4. Appendix 4: Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

- Summary tables of AEs, provided as datasets which will be converted to an XML files.
- An AE is considered 'Serious' whether or not it is a TEAE.
- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each Serious AE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term
    - the number of events experienced.

6.5. Appendix 5: COVID-19 Pandemic Impact

This section lists additional statistical analyses that may be performed at the database lock to assess the impact of COVID-19 if the data warrant.

6.5.1. General Consideration

Percentage and count of randomized participants who followed the COVID-19 mitigation plan may be summarized by treatment group. This includes, but is not limited to, participants rescreened, procedures conducted via remote visit or mobile home health visit, visits occurred using the extended visit windows, alternative way of investigator product shipment/dispensing, use of a local lab, etc. A listing of randomized participants who followed the COVID-19 mitigation plan may be provided. Similar analyses may be provided by country and by treatment group.

Percentage and count of randomized participants whose study visits were impacted by COVID-19 pandemic may also be summarized. A listing may be provided.

6.5.2. Exposure

A listing of randomized participants who had study drug temporarily interrupted due to COVID-19 pandemic may be provided.

6.5.3. Protocol Deviation

Percentage and count of randomized participants having important protocol deviations related to COVID-19 pandemic will be summarized by treatment.

Percentage and count of randomized participants with protocol deviations related to COVID-19 pandemic may also be summarized by treatment.

A listing of all randomized participants who had important protocol deviations due to COVID-19 pandemic may be provided.

6.5.4. Patient Disposition

A summary table for all randomized participants that discontinue study or study treatment due to COVID-19 pandemic will be provided by treatment.

A listing of randomized participants who discontinued the study or study treatment due to COVID-19 pandemic will be provided.

6.5.5. Adverse Events

A listing of all enrolled participants who had COVID-19 infection, including death due to COVID-19, during the post-randomization period will be provided. A summary table may be provided if deemed necessary.

6.5.6. Major Depressive Disorder/Suicidal Ideation

The counts and percentages of participants with TEAEs for major depression may be summarized by treatment group using MedDRA PT nested within SMQ by COVID-19 subgroup (that is, participants without impact versus with impact) for SS group. Detailed searching criteria can be found in Appendix 6 (Section 6.6).

A participant is defined as impacted by COVID-19 if either one of the following is satisfied:

- no COVID-19 illness, but impacted by quarantine and travel restrictions, clinics closing, visits being canceled, delay or non-delivery of the investigational product, virtual visits, etc.

OR

- with COVID-19 illness.

The suicidal ideation and behavior solicited through C-SSRS may be summarized by treatment group by COVID-19 subgroup (that is, participants without impact vs with impact) for SS group.

6.5.7. Local Lab

Local lab performed due to exceptional circumstances will not be brought into the Lilly database at the time of final database lock per data collection system, even though local laboratory is one of the options in exceptional circumstances. Therefore, this section is not applicable for analysis purpose.

6.5.8. Missing Data Due to Exceptional Circumstances

For the primary endpoints, missing data due to exceptional circumstances will be handled as described in Section 4.3.3.2. In addition, a summary table for participants whose primary measurements were impacted by COVID-19 (including missing, collected using alternative options) may be provided. A listing of participants whose primary measurements were impacted by COVID-19 (including missing, collected using alternative options) may be provided if deemed necessary.

6.6. Appendix 6: Searching Criteria for Special Safety Topics

Abuse Liability

To identify AE terms suggestive of potential abuse liability, narrow terms from SMQ of Drug abuse and dependence (20000101) will be used.

Acute Gallbladder Disease

All biliary colic, cholecystitis, or other suspected events related to gallbladder disease will be identified using the MedDRA PTs in any of the following:

- Narrow PTs in Gallbladder related disorders SMQ (20000124)
- Narrow PTs in Biliary tract disorders SMQ (20000125)
- Narrow PTs in Gallstone related disorders SMQ (20000127).

Amputation/Peripheral Revascularization

Amputations/peripheral revascularization events will be identified using the following MedDRA PTs:

- Amputation
- Peripheral revascularization.

C-cell Hyperplasia and Thyroid Malignancies

Thyroid malignancies and C-Cell hyperplasia will be identified using MedDRA HLT for Thyroid neoplasms and PT for thyroid C-cell hyperplasia.

Hepatic Events

Treatment-emergent potentially drug-related hepatic events will be identified using the MedDRA PTs contained in any of the following:

- Broad and narrow terms in the Liver related investigations, signs and symptoms SMQ (20000008)
- Broad and narrow terms in the Cholestasis and jaundice of hepatic origin SMQ (20000009)
- Broad and narrow terms in the Hepatitis non-infections SMQ (20000010)
- Broad and narrow terms in the Hepatic failure, fibrosis and cirrhosis and other liver damage SMQ (20000013)

- Narrow terms in the Liver-related coagulation and bleeding disturbances SMQ (20000015).

Hypersensitivity Reactions

Treatment-emergent hypersensitivity reactions will be identified based on the following:

- Narrow and algorithm terms in Anaphylactic reaction SMQ (20000021)
- Narrow terms in Angioedema SMQ (20000024)
- Narrow terms in Severe cutaneous adverse reactions SMQ (20000020)
- Narrow terms in Hypersensitivity SMQ (20000214).
- Narrow terms in Vasculitis SMQ (20000174).

For the Anaphylactic reaction SMQ, each term is classified by scope (Narrow, Broad) and by category (A, B, C, D). All Narrow terms are category A, and all Broad terms are category B, C, or D. In addition to the usual Narrow and Broad searches, the SMQ defines an algorithm to further refine the cases of interest. For Time Period A of potential immediate hypersensitivity analysis, the Anaphylactic reaction SMQ algorithm will be included. The algorithm is based upon events that occur within Time Period A. The counts and percentages of participants who experienced a TEAE for the following will be analyzed for each of Time Period A and Time Period B (potential non-immediate hypersensitivity analysis):

- any narrow term from any one of the 5 SMQs indicated above (that is, combined search across narrow of all 5 SMQs)
- any narrow scope term within each SMQ, separately (that is, narrow SMQ search). For analysis in Time Period A, any term from Anaphylactic reaction SMQ algorithm.

Injection Site Reactions

The ISR AE will be identified using the MedDRA PT in any of the following:

- HLT of Injection site reaction
- HLT of Administration site reaction
- HLT of Infusion Site Reactions

Pancreatitis Events

Determination of investigator-reported events will be through the “Acute pancreatitis” Standardized MedDRA Query (SMQ) (20000022, narrow scope) and a “Pancreatitis Chronic” PT search of the AE database, while adjudication-confirmed pancreatitis are found from adjudication forms.

Malignancies

The malignancy events will be identified using the MedDRA PT contained in Malignant tumours SMQ (20000194) narrow scope or Tumours of unspecified malignancy SMQ (20000195) narrow scope.

Arrhythmias and Cardiac Conduction Disorders

The arrhythmias and cardiac conduction disorders events will be identified using the MedDRA PT contained in any of the following SMQs:

1) Arrhythmias:
    - For symptoms: Arrhythmia related investigations, signs and symptoms SMQ (20000051), narrow and broad terms
    - For supraventricular arrhythmias: In Cardiac arrhythmia SMQ, under tachyarrhythmia sub SMQ
        - Supraventricular tachyarrhythmia SMQ (20000057), broad and narrow terms
        - Tachyarrhythmia terms, nonspecific SMQ (20000164), narrow terms only; and
        - Ventricular tachyarrhythmia SMQ (20000058), narrow terms only.
2) Cardiac Conduction Disorders
    - Conduction defects SMQ (20000056), narrow terms only; and
    - Cardiac conduction disorders High Level Term (HLT; 10000032), all PTs.

Injection Site Reactions

Treatment emergent injection site reaction will be identified using the MedDRA PT in any of the following:

- MedDRA HLT of Injection site reaction
- HLT of Administration site reaction
- HLT of Infusion Site Reactions

Major Depressive Disorder/Suicidal Ideation

The major depressive disorder/suicidal ideation or behavior will be captured as AESI. Adverse events will be searched using MedDRA PT terms. The PTs from the Depression and suicide/self-injury SMQ as defined in MedDRA (SMQs: 20000037 [Suicide/self-injury] and 20000167 [Depression (excl suicide and self-injury)]) will be summarized.

Amputation/Peripheral Revascularization

Amputation/Peripheral revascularization will be considered as AESIs. Participants with amputations/peripheral revascularization will be searched using the following MedDRA PTs:

- Amputation and
- Peripheral revascularization.

Metabolic Acidosis, Including Diabetic Ketoacidosis

Metabolic acidosis including diabetic ketoacidosis will be searched using the following MedDRA version at the time of database lock PTs :

- Diabetic ketoacidosis
- Ketoacidosis
- Euglycaemic diabetic ketoacidosis
- Ketonuria
- Diabetic ketosis
- Diabetic ketoacidotic hyperglycaemic coma
- Ketosis
- Lactic acidosis
- Urine ketone body present
- Blood ketone body
- Blood ketone body increased
- Urine ketone body, and
- Blood ketone body present.

7. References

Araki E, Goto A, Kondo T, et al. Japanese Clinical Practice Guideline for Diabetes 2019. *J Diabetes Investig*. 2020;11(4):1020-1076.

Rubin DB. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons Inc.; 1987.

EuroQoL Group. EQ-5D-5L User guide: basic information on how to use the EQ-5D-5L instrument. Version 2.1. April 2015. Accessed January 22, 2019. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf.

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.

Ikeda S, Shiroiwa T, Igarashi I, et al. Developing a Japanese version of the EQ-5D-5L value set (in Japanese; abstract in English). *J Natl Inst Public Health*. 2015;64:47-55.

[JASSO] Japan Society for the Study of Obesity. Guideline for the management of obesity disease 2016. Tokyo: Life Science Publishing. 2016. (in Japanese)

Kolotkin RL, Crosby RD, Williams GR. Health-related quality of life varies among obese subgroups. *Obes Res*. 2002;10(8):748-756. https://doi.org/10.1038/oby.2002.102

Tokushige K, Ikejima K, Ono M, Eguchi Y, Kamada Y, Itoh Y, Akuta N, Yoneda M, Iwasa M, Yoneda M, Otsuka M, Tamaki N, Kogiso T, Miwa H, Chayama K, Enomoto N, Shimosegawa T, Takehara T, Koike K. Evidence-based clinical practice guidelines for nonalcoholic fatty liver disease/nonalcoholic steatohepatitis 2020. *J Gastroenterol*. 2021;56(11):951-963.

Signature Page for VV-CLIN-066996 v1.0

Approval	PPD 11-Jul-2023 00:42:26 GMT+0000

Signature Page for VV-CLIN-066996 v1.0