Statistical Analysis Plan: I8F-MC-GPHL (Version 2)

Efficacy and Safety of Tirzepatide Once Weekly in Participants with Type 2 Diabetes Who Have Obesity or Are Overweight: A Randomized, Double-Blind, Placebo-Controlled Trial (SURMOUNT-2)

NCT04657003

Approval Date: 13 Jan 2023

Title Page

Protocol Title: Efficacy and Safety of Tirzepatide Once Weekly in Participants with Type 2 Diabetes Who Have Obesity or Are Overweight: A Randomized, Double-Blind, Placebo-Controlled Trial (SURMOUNT-2)

Protocol Number: I8F-MC-GPHL

Compound Number: LY3298176

Short Title: Effect of Tirzepatide versus Placebo in Participants with Type 2 Diabetes Who Have Obesity or Are Overweight (SURMOUNT-2)

Acronym: SURMOUNT-2

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

IND 139721

Document ID: VV-CLIN-076568

Table of Contents

Abbreviations and Definitions

Term	Definition
ADA	anti-drug antibodies
AE	adverse event
AESI	adverse events of special interest
AHM	antihyperglycemic mediation
ALT	alanine transaminase
ANCOVA	analysis of covariance
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
eGFR	estimated glomerular filtration rate
ED	early discontinuation
FAS	Full Analysis Set
FDA	Food and Drug Administration
GI	gastrointestinal

Term	Definition
GIP	glucose-dependent insulinotropic polypeptide
GLP-1	glucagon-like peptide-1
GIPR	glucose-dependent insulinotropic polypeptide receptor
GLP-1R	glucagon-like peptide-1 receptor
HbA1c	hemoglobin A1c
HDL	high-density lipoprotein
HLT	High Level Term
ICE	intercurrent event
ICH	International Council for Harmonisation
ISR	injection-site reaction
IWQOL-Lite-CT	Impact of Weight on Quality of Life-Lite-Clinical Trials Version
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
MRD	minimum required dilution
NAb	neutralizing antibodies
PD	pharmacoydynamic
PG	plasma glucose
PHQ-9	Patient Health Questionnaire-9
PGIS	Patient Global Impression of Status for Physical Activity
PK	pharmacokinetic
PT	Preferred Term

Term	Definition
QTcF	Fredericia's corrected QT interval
QW	once weekly
REML	restricted maximum likelihood
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	standard deviation
SI	Système International
SMQ	Standardised MedDRA Query
SOC	System Organ Class
SS	Safety Analysis Set
T2DM	type 2 diabetes mellitus
TBL	total bilirubin
TE	treatment-emergent
TEAE	treatment-emergent adverse event
UACR	urine albumin-to-creatinine ratio
ULN	upper limit of normal
VAS	Visual Analog Scale

Version History

This SAP for Study I8F-MC-GPHL (GPHL) is based on the protocol dated 23 Nov 2020.

SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	9-Sep-2022	Not Applicable	Original version
2	See date on page 1	See below summary	See below summary

The key changes of SAP Version 2 are summarized below:

1. Added mean change in EQ-5D-5L utility score and VAS score to "Additional Secondary objectives, by dose analysis" in Table GPHL.1.1 and Section 4.9 to align with analyses described in Table GPHL.4.3 (see Section 1.1.1).
2. In response to US FDA feedback, clarified that the population-level summary for the "difference in response percentages between treatment conditions" is assessed by odds ratio (see Section 1.1.2.1).
3. Clarified that placebo multiple imputation will be used for missing data imputation in Category 1 for treatment-regimen estimand if there are not enough retrieved dropouts to impute missing data in Category 2 (see Section 4.3.2.3).
4. In response to US FDA feedback, added sensitivity analyses to evaluate the robustness of the primary efficacy results using different missing data imputation methods (see Section 4.3.2.4).
5. Updated to use "baseline HbA1c value" rather than "baseline HbA1c group (≤8.5%, >8.5%)" as a covariate in analysis models for efficacy measures related to HbA1c to better adjust for baseline HbA1c information (see Section 4.4.1).
6. Added exploratory endpoints to assess risk difference among each tirzepatide dose and placebo arm for an unconditional effect in proportions of participants achieving body weight reduction targets (see Section 4.5).

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

Changes to the protocol-planned objectives and endpoints are described in Section 4.9.

Table GPHL.1.1. Objectives and Endpoints

Objectives	Endpoints
Primary	
To demonstrate that tirzepatide 10 mg and/or 15mg QW is superior to placebo at 72 weeks for: • Percent change in body weight AND • Proportion of participants with ≥5% body weight reduction	• Mean percent change in body weight from randomization • Percentage of study participants who achieve ≥5% body weight reduction from randomization
Key Secondary (controlled for type I error), by dose analysis	
For QW tirzepatide 10 mg and/or 15 mg doses, to demonstrate superiority to placebo in change from randomization for the following (at 72 weeks): • Body weight	• Percentage of participants who achieve: o ≥10% body weight reduction o ≥15% body weight reduction o ≥20% body weight reduction
• Glycemic control	• Mean change in HbA1c (%) • Percentage of participants who achieve: o HbA1c <7% o HbA1c ≤6.5% o HbA1c <5.7% • Mean change in fasting glucose (mg/dL)
• Waist circumference	• Mean change in waist circumference (cm)

Objectives	Endpoints
Key secondary (controlled for type I error), pooled dose analysis	
For QW tirzepatide (all doses combined), to demonstrate superiority to placebo in change from randomization for the following (at 72 weeks):	
• Lipid parameters	• Mean change in fasting o Triglycerides (mg/dL) o HDL-cholesterol (mg/dL) o Non-HDL-cholesterol (mg/dL)
• SBP	• Mean change in SBP (mmHg)
Additional Secondary, by dose analysis	
For QW tirzepatide 10 mg and/or 15 mg doses, to demonstrate superiority to placebo in change from randomization for the following (at 72 weeks):	
• Body weight	• Mean change in absolute body weight (kg) • Mean change in BMI (kg/m^2)
• 7-point self-monitored blood glucose profiles	• Mean change in o each of the 7-point measurements o 2-hour morning, midday and evening meal excursions, and all meals 2-hour excursion o mean of all 7-point measurements, mean of all pre-meal measurements, and mean of all 2-hour postprandial measurements
• Insulin	• Mean change in fasting insulin (pmol/L)
• Patient-Reported Outcomes	• Mean change in o SF-36v2 acute form Physical Functioning domain score o IWQOL-Lite-CT Physical Function composite score o EQ-5D-5L utility score and VAS score

Objectives	Endpoints
Additional Secondary, pooled analysis	
For QW tirzepatide (all doses combined), to demonstrate superiority to placebo in change from randomization for the following (at 72 weeks):	
• Lipid parameters	• Mean change in fasting o Total cholesterol (mg/dL) o LDL-cholesterol (mg/dL) o VLDL-cholesterol (mg/dL) o Free fatty acids (mg/dL)
• DBP	• Mean change in DBP (mmHg)

Abbreviations: BMI = body mass index; DBP = diastolic blood pressure; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite-Clinical Trials Version; LDL = low-density lipoprotein; QW = once weekly; SBP = systolic blood pressure; SF-36v2 acute form = Short Form-36 Health Survey version 2 acute form; VAS = Visual Analog Scale; VLDL = very low-density lipoprotein.

1.1.2. Estimands

1.1.2.1. Estimands for Co-primary Endpoints

The primary clinical question of interest is: What is the intervention difference between QW tirzepatide 10 mg and/or 15 mg and placebo in mean percent change in body weight and proportion of participants who achieve ≥5% body weight reduction from randomization to 72 weeks in participants with T2DM, who have obesity (BMI ≥30 kg/m^2) or overweight (BMI ≥27 kg/m^2)?

For primary disclosure of Study GPHL including submission to the US Food and Drug Administration (FDA)

The estimand is described by the following attributes:

- Population: all randomized participants who received at least 1 dose of study drug.
- Co-primary endpoints: 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 versus placebo, regardless of study drug adherence.
- Handling of intercurrent events: the intercurrent event leading to treatment discontinuation for any reason is addressed by the treatment condition of interest attribute and handled by treatment policy strategy as described in ICH E9 (R1) Addendum (ICH 2019). Further details can be found in Section 4.3.2.1.

- Population-level summary: difference in mean changes between treatment conditions, AND difference in response percentage between treatment conditions as assessed by odds ratio.

Rationale for estimand: It aims to reflect treatment effect among participants with T2DM and with obesity or overweight, irrespective to the compliance to the planned course of treatment.

This *de facto estimand* is referred to as the "treatment regimen" estimand in the latter sections of this document, which is equivalent to the "hybrid" estimand as stated in the protocol.

For other purposes

The estimand is described by the following attributes:

- Population: all randomized participants who received at least 1 dose of study drug.
- Co-primary endpoints: 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 versus placebo, excluding data after discontinuation of study drug.
- Handling of intercurrent events: the intercurrent event leading to treatment discontinuation for any reason is addressed by the treatment condition of interest attribute and handled by the hypothetical strategy as described in ICH E9 (R1) Addendum (ICH 2019).
- Population-level summary: difference in mean changes between treatment conditions, AND difference in response percentage between treatment conditions as assessed by odds ratio.

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.

We will refer to this estimand as "efficacy" estimand in the latter sections.

1.1.2.2. Estimands for Secondary Endpoints

Table GPHL.1.2. Attribute Description of Estimands for Secondary Endpoints

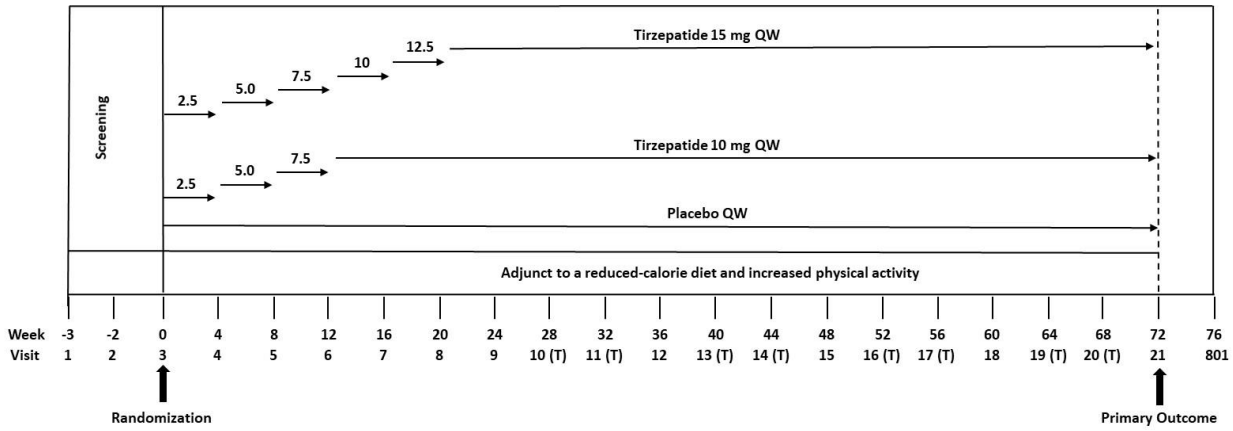
Estimand for Secondary Endpoints	Population	Endpoint	Treatment condition of interest	Handling of intercurrent events	Population-level summary
Key secondary related to mean change	Same as estimands for co-primary endpoints	As described in Table GPHL.1.1	For primary disclosure of study GPHL including submission to the US FDA, follow the same manner of the treatment regimen estimand in Section 1.1.2.1; for all other purposes, follow the same manner of the 'efficacy' estimand in Section 1.1.2.1	For primary disclosure of study GPHL including submission to the US FDA, follow the same manner of the treatment regimen estimand in Section 1.1.2.1; for all other purposes, follow the same manner of the 'efficacy' estimand in Section 1.1.2.1	Difference in mean changes between treatment conditions
Key secondary related to percentage of participants meeting certain criteria					Difference in percentage of participants between treatment conditions as assessed by odds ratio
Other secondary related to mean change	Same as estimands for co-primary endpoints	As described in Table GPHL.1.1	Follow the same manner of the 'efficacy' estimand in Section 1.1.2.1	Follow the same manner of the 'efficacy' estimand in Section 1.1.2.1	Difference in mean changes between treatment conditions

Abbreviation: FDA = Food and Drug Administration.

Unless otherwise specified, the attributes of estimands supporting exploratory objectives will follow the same manner of the efficacy estimand in Section 1.1.2.1.

1.2. Study Design

Study I8F-MC-GPHL (GPHL; SURMOUNT-2) is a Phase 3, multicenter, randomized, placebo-controlled, double-blinded study of the safety and efficacy of 10 mg and 15 mg tirzepatide QW, compared with placebo, when used in conjunction with a reduced-calorie diet and increased physical activity for weight management, in participants who have T2DM, and have obesity (BMI ≥30 kg/m^2) or overweight (BMI ≥27 kg/m^2). All participants will be randomized to 72 weeks of treatment to study the effects on body weight reduction.



Abbreviations: QW = once weekly; T = telephone visit.
Note: All participants will be randomized to 72 weeks of treatment to study the effects on body weight reduction. The safety follow-up visit will occur 4 weeks after participants discontinue or complete the treatment.

Figure GPHL.1.1. Illustration of study design for Clinical Protocol I8F-MC-GPHL.

The details about the overview of study periods and study visits can be found in Study GPHL protocol Section 4. The detail of the unique visits not displayed in Figure GPHL.1.1 is provided below.

Visit 99

Visit 99 is only applicable to participants who discontinue the study treatment prematurely prior to Visit 21 (Week 72). Participants will be asked to return for Visit 99 at 72 weeks ± 7 days after randomization. This visit is critical to ensure complete data collection for the primary weight-reduction endpoint.

Participants should attend this visit in the fasting state. Procedures to be completed include

- measurement of weight and waist circumference
- concomitant medications
- assessment of AEs, and
- completion of the mental health questionnaires (after the AE assessment).

Early Discontinuation of Treatment Visit

Participants unable or unwilling to continue the study treatment for any reason will perform early discontinuation of treatment visit at the visit when the participant informs the site about the study treatment discontinuation.

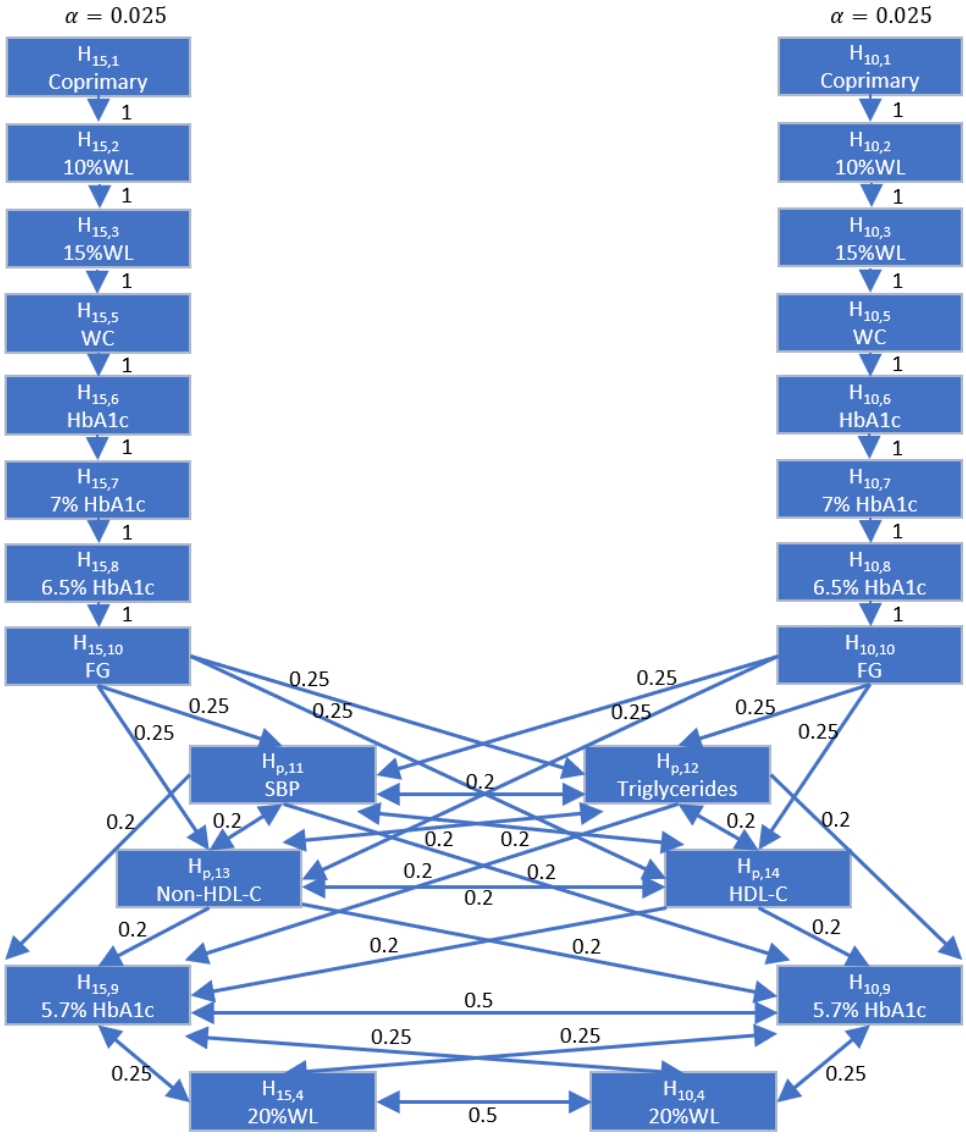
2. Statistical Hypotheses

The alternative hypothesis for the primary objective and key secondary objectives are the following:

- $H_{15,1}$ and $H_{10,1}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for mean percent change in body weight from randomization and percentage of participants who achieve ≥5% body weight reduction at 72 weeks.
- $H_{15,2}$ and $H_{10,2}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for percentage of participants who achieve ≥10% body weight reduction from randomization at 72 weeks
- $H_{15,3}$ and $H_{10,3}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for percentage of participants who achieve ≥15% body weight reduction from randomization at 72 weeks
- $H_{15,4}$ and $H_{10,4}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for percentage of participants who achieve ≥20% body weight reduction from randomization at 72 weeks
- $H_{15,5}$ and $H_{10,5}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for change from randomization in waist circumference (cm) at 72 weeks
- $H_{15,6}$ and $H_{10,6}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for change from randomization in HbA1c (%) at 72 weeks
- $H_{15,7}$ and $H_{10,7}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for percentage of participants who achieve HbA1c <7% at 72 weeks
- $H_{15,8}$ and $H_{10,8}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for percentage of participants who achieve HbA1c ≤6.5% at 72 weeks
- $H_{15,9}$ and $H_{10,9}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for percentage of participants who achieve HbA1c <5.7% at 72 weeks
- $H_{15,10}$ and $H_{10,10}$: superiority of tirzepatide 10 mg and/or 15 mg versus placebo for change from randomization in fasting glucose at 72 weeks
- $H_{p,11}$: superiority of tirzepatide (all doses combined) versus placebo for change in SBP (mmHg) from randomization at 72 weeks
- $H_{p,12}$: superiority of tirzepatide (all doses combined) versus placebo for change in triglycerides (mg/dL) from randomization at 72 weeks
- $H_{p,13}$: superiority of tirzepatide (all doses combined) versus placebo for change in non-HDL-cholesterol (mg/dL) from randomization at 72 weeks
- $H_{p,14}$: superiority of tirzepatide (all doses combined) versus placebo for change in HDL-cholesterol (mg/dL) from randomization at 72 weeks

2.1. Multiplicity Adjustment

For primary and key secondary efficacy objectives, the details of type I error control strategy are illustrated in Figure GPHL.2.1. Both efficacy estimand and treatment regimen estimand will be used to assess these objectives. As they are intended for different purposes, no multiplicity adjustments will be made for conducting separate analyses relative to the efficacy and treatment regimen estimands. In addition, no multiplicity adjustments will be made for evaluating other secondary and exploratory objectives, or for safety assessments.



Abbreviations: FG = fasting glucose; H = hypothesis; HbA1c = hemoglobin A1c; HDL-C = high-density lipoprotein cholesterol; Non-HDL-C = non high-density lipoprotein cholesterol; p = pooled dose; SBP = systolic blood pressure; WC = waist circumference; WL = weight loss.

Figure GPHL.2.1. Type 1 error control strategy for primary and key secondary efficacy endpoints.

3. Analysis Sets

For purposes of analysis, the following populations are defined:

Table GPHL.3.1. Description of Analysis Datasets

Population	Description
Entered	All participants who sign informed consent
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.
Efficacy Analysis Set (EAS)	For glycemic control related endpoints: Data obtained during treatment period from mITT, excluding data after initiation of rescue antihyperglycemic medication or premature discontinuation of study drug (last dose date + 7 days). For other endpoints: Data obtained during treatment period from mITT, excluding data after premature discontinuation of study drug (last dose date + 7 days).
Full Analysis Set (FAS)	Data obtained during treatment period from mITT, regardless of adherence to study drug or initiation of rescue antihyperglycemic medication.
Safety Analysis Set (SS)	Data obtained during the treatment period plus safety follow-up period from mITT, regardless of adherence to study drug or initiation of rescue antihyperglycemic medication.

Unless otherwise specified, for analyses guided by the treatment regimen estimand, the FAS will be used; for analyses guided by the efficacy estimand, the EAS will be used; for safety analysis, the 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. Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (eg, few events to justify conducting an analysis). Additional analyses of the data may be conducted as deemed appropriate.

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

Unless specified otherwise, efficacy and safety will be assessed using the mITT population, and data will be analyzed based on the assigned treatment (ie, not the actual treatment received by the participant).

Summary descriptive 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 either ANCOVA or a MMRM.

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 hazards 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 covariate. Otherwise, Fisher's exact test will be used to examine the treatment difference.

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

Unless specified otherwise, baseline will be defined as the last available nonmissing measurement during Visit 1 to Visit 3. For the safety-related parameters, the definition of baseline and postbaseline are specified in Table GPHL.4.1.

Table GPHL.4.1. Baseline and Postbaseline Definitions for Safety Groups

Analysis Set	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 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 measurements after Visit 3. 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 (1.2). Only scheduled visits will be included. The ED visits are considered scheduled visits.

Abbreviations: ECGs = electrocardiogram; ED = early discontinuation; SS = Safety Analysis Set.

[a] Immunogenicity related analysis is specified in Section 4.6.5.5.

For the primary and key secondary efficacy endpoint analyses subject to type 1 error rate control, data for participants with missing values at the 72-week visit will be imputed based on the method described in Section 4.3.2.3. 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 early discontinuation 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 and placebo. Since the trial is not adequately powered to detect differences 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 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

A listing of study disposition for all randomized participants will be provided at the final database lock. Summaries of study disposition and study drug disposition for all randomized participants will be provided by planned study treatment at the final database lock.

4.3. Primary Endpoints Analysis

For submission of Study GPHL to the US FDA to support the registration of tirzepatide for chronic weight management, both primary efficacy assessments will be guided by the treatment regimen estimand conducted using the FAS. Assessment of the primary objectives will be conducted with hybrid imputation of missing data (see Section 4.3.2.3).

For other regulatory agencies, publications and other purposes, the assessment of efficacy objectives may be guided by the efficacy estimand using the EAS.

4.3.1. Definition of Endpoint(s)

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 is defined as:

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

4.3.2. Main Analytical Approach

4.3.2.1. The Analysis Related to the Treatment Regimen Estimand

The analysis related to 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 at 72 weeks. A logistic regression model will be used for the analysis of the percentage of participants achieving at least 5% body weight reduction obtained at the 72-week visit. Both models will include terms of treatment group, country/pooled country, sex, and type of AHM (significant weight gain, significant weight loss, weight neutral and others) used at randomization as fixed effects and baseline body weight as a covariate. The ANCOVA analysis will be conducted with hybrid imputation of missing body weight at 72 weeks (see Section 4.3.2.3 for details) and statistical inference over multiple imputation of missing data guided by Rubin (1987). As for the logistic regression, missing body weight data at 72 weeks will be imputed first based on imputation method specified in Section 4.3.2.3, then the continuous measurements will be categorized into status of achieving at least 5% body weight reduction (Yes or No).

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

4.3.2.2. The Analysis Related to the Efficacy Estimand

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

For the mean percent body weight change from randomization, a MMRM will be conducted. 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 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 achieving at least 5% body weight reduction (Yes or No) at 72-week visit, with missing value imputed using the predicted value from MMRM analysis followed by dichotomization.

For MMRM the independent variables of analysis model are treatment group, visit, treatment-by-visit interaction, stratification factors (type of AHM used at randomization, sex, and country/pooled country) 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, and
- compound symmetry.

The first covariance structure that converges will be used.

With the aid of the MMRM analysis, 2-sided 95% CIs for mean percent change in body weight from randomization to the 72-week visit for tirzepatide 10 mg and 15 mg 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.

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 and placebo, as well as tirzepatide 15 mg and placebo will be derived.

4.3.2.3. Methods for Hybrid Imputations

For efficacy analyses relative to “treatment-regimen” estimand, missing values resulting from ICEs will be handled as follows:

- Category 1: for missing data solely due to exceptional circumstances, such as pandemic or natural disasters (after other reasons for missing data are ruled out), the analysis will consider 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 ICEs, missing data 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 the study drug. In cases where there are not enough retrieved dropouts to provide a reliable imputation model (eg, 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.

In cases where placebo multiple imputation method is used for missing data imputation in Category 2 due to not enough retrieved dropouts, the missing data in Category 1 will also be imputed using all nonmissing data of the primary outcome measurement from the placebo group.

4.3.2.4. Sensitivity Analysis Related to Treatment Regimen Estimand

For submission of Study GPHL to the US FDA to support the registration of tirzepatide for chronic weight management, additional sensitivity analyses of the primary efficacy outcomes will be conducted using the FAS and guided by the "treatment-regimen" estimand, which represents the efficacy irrespective of adherence to study drug. This assessment will analyze percent change in body weight obtained at the 72-week visit using an ANCOVA and the percentage of participants achieving at least 5% body weight reduction obtained at the 72-week visit using a logistic regression model. The terms for both models will be the same as specified in Section 4.3.2.1. These analyses are intended to assess the robustness of primary efficacy results using different missing data imputation methods:

- Retrieved dropout multiple imputation: Missing values of change in body weight at the 72-week visit will be imputed based on observed body weight change from baseline values at the visit 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), the placebo multiple imputation method (described below) will be used.
- Placebo multiple imputation: Missing values of change in body weight at the 72-week visit will be imputed based on observed body weight change from baseline values at the visit from participants in the placebo treatment group.
- Return to baseline imputation: Missing values of body weight at the 72-week visit will be imputed using the return-to-baseline multiple imputation method to account for within subject variability (Qu et al. 2022).

4.4. Secondary Endpoint(s)/Estimands(s)] Analysis

4.4.1. Key Secondary Analyses Subject to Type 1 Error Rate Control

Table GPHL.4.2. Key Secondary Measures (Controlled for Type 1 Error)

Objectives	Relative to the efficacy measure:	Analysis conducted in a manner similar to	Additional Information
Tirzepatide 10 mg and/or 15 mg is superior to placebo:	Percentage of participants achieving body weight reduction ≥10%, ≥15% and ≥20% at 72 weeks	logistic model in Section 4.3.2.1 for treatment-regimen estimand and logistic model in Section 4.3.2.2 for efficacy estimand	
	Mean change in HbA1c (%) from randomization at 72 weeks	ANCOVA model in Section 4.3.2.1 for treatment-regimen estimand and MMRM model in Section 4.3.2.2 for efficacy estimand, with baseline HbA1c value to replace baseline body weight as an additional covariate.	LSM estimates will be plotted by treatment through 72 weeks.
	Percentage of participants achieving target value of HbA1c <7%, ≤6.5% and <5.7% at 72 weeks	logistic model in Section 4.3.2.1 for treatment-regimen estimand and logistic model in Section 4.3.2.2 for efficacy estimand, with baseline HbA1c value to replace baseline body weight as an additional covariate.	
	Mean change in waist circumference (cm) from randomization at 72 weeks	ANCOVA model in Section 4.3.2.1 for treatment-regimen estimand and MMRM model in Section 4.3.2.2 for efficacy estimand, with baseline waist circumference to replace baseline body weight as an additional covariate.	LSM estimates will be plotted by treatment through 72 weeks.
	Mean change in fasting glucose (mg/dL) from randomization at 72 weeks	ANCOVA model in Section 4.3.2.1 for treatment-regimen estimand and MMRM model in Section 4.3.2.2 for efficacy estimand	Baseline fasting glucose will be used a covariate and baseline HbA1c group (≤8.5%, >8.5%) will be used as a fixed effect. LSM estimates will be plotted by treatment through 72 weeks.
Tirzepatide (all doses combined) is superior to placebo:	Mean change in triglycerides, HDL-cholesterol, and non-HDL-cholesterol from randomization at 72 weeks	ANCOVA model in Section 4.3.2.1 for treatment-regimen estimand and MMRM model in Section 4.3.2.2 for efficacy estimand, with baseline lipids value in place of baseline body weight as a covariate.	Estimated means will be plotted by treatment through 72 weeks. Log transformation will be adopted for lipid parameters.

Objectives	Relative to the efficacy measure:	Analysis conducted in a manner similar to	Additional Information
	Mean change in SBP (mmHg) from randomization at 72 weeks	ANCOVA model in Section 4.3.2.1 for treatment-regimen estimand and MMRM model in Section 4.3.2.2 for efficacy estimand, with baseline SBP in place of baseline body weight as a covariate.	LSM estimates will be plotted by treatment through 72-weeks.

Abbreviations: ANCOVA = analysis of covariance; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; LSM = least squares mean; MMRM = mixed model for repeated measures; SBP = systolic blood pressure.

Decision regarding reject/accept of the statistical hypothesis for each objective will be guided by the 2-sided p-values, and details are included in Section 2.1.

For submission of Study GPHL to the US FDA to support the registration of tirzepatide for chronic weight management, the key secondary efficacy assessment will be guided by the treatment regimen estimand, conducted using the same population as for the primary analysis. Assessment of the key secondary objectives will be conducted with hybrid imputation of missing data (see Section 4.3.2.3).

For other regulatory agencies, publications, and other purposes, the assessment of key secondary efficacy objectives may be guided by the efficacy estimand using the same population as for primary analysis.

4.4.2. Supportive Secondary Endpoints

Unless otherwise specified, other secondary and exploratory efficacy analyses (not controlled for type 1 error) will be conducted using the EAS. Missing data will be imputed using LOCF or handled through MMRM without utilizing hybrid imputation technique.

Analyses for labs including 7-point self-monitored blood glucose, fasting insulin, and lipid parameters will be performed for both SI unit and CN unit.

4.4.2.1. Additional Secondary Efficacy Analyses

Table GPHL.4.3. Secondary Measures (Not Controlled for Type 1 Error)

Objective	Relative to the efficacy measure:	Analysis conducted in a manner similar to:	Additional Information
compare tirzepatide 10 mg and/or 15 mg QW with placebo at 72 weeks:	Mean change in body weight (kg) from randomization	MMRM model in Section 4.3.2.2	LSM estimates through 72 weeks will be plotted by study treatment.
	Mean change in BMI (kg/m^2) from randomization	MMRM model in Section 4.3.2.2	Baseline BMI (kg/m^2) will be used as a covariate. LSM estimates through 72 weeks will be plotted by study treatment.

Objective	Relative to the efficacy measure:	Analysis conducted in a manner similar to:	Additional Information
	Mean change in 7-point SMBG from randomization	MMRM model in Section 4.3.2.2	Baseline SMBG value will be used a covariate and baseline HbA1c group (≤8.5%, >8.5%) will be used as a fixed effect. In addition to the analyses on each of the 7-points, similar analyses will be done for the 2-hour morning, midday, and evening meal excursions, the mean of all meals 2-hour excursion, the mean of all 7-point measurements, the mean of all pre-meal measurements, and the mean of all 2-hour postprandial measurements.
	Mean change in fasting insulin from randomization	MMRM model in Section 4.3.2.2	Baseline fasting insulin will be used as a covariate. Estimated means through 72 weeks will be plotted by study treatment. Log transformation will be adopted for fasting insulin.
	Mean change in IWQOL-Lite-CT Physical Function composite (PF) score from randomization	ANCOVA model	Terms of treatment, stratification factors, and baseline IWQOL-Lite CT PF score will be used as a covariate. Missing data will be imputed using LOCF method.
	Mean change in SF-36v2 acute form physical functioning domain (PF) score from randomization	ANCOVA model	Same as above but using baseline SF-36 v2 acute PF score as a covariate.
	Mean change in EQ-5D-5L utility score and VAS score from randomization	ANCOVA model	Same as above but using baseline EQ-5D-5L utility score and VAS score as a covariate, respectively.
compare tirzepatide QW (all doses combined) with placebo at 72 weeks	Mean change in DBP (mmHg)	MMRM model in Section 4.3.2.2	All tirzepatide doses will be pooled together. LSM estimates through 72 weeks will be plotted by study treatment.

Objective	Relative to the efficacy measure:	Analysis conducted in a manner similar to:	Additional Information
	Mean change in LDL-C, VLDL-C, total-cholesterol, and free fatty acids from randomization	MMRM model in Section 4.3.2.2	All tirzepatide doses will be pooled together. Estimated means through 72 weeks will be plotted by study treatment. Log transformation will be adopted for lipids parameters.

Abbreviations: ANCOVA = analysis of covariance; BMI = body mass index; DBP = diastolic blood pressure; EQ-5D-5L = EuroQol-5 Dimension 5 Level; HbA1c = hemoglobin A1c; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite-Clinical Trials Version; LDL-C = low-density lipoprotein cholesterol; LOCF = last observation carried forward; LSM = least squares mean; MMRM = mixed model for repeated measures; QW = once-weekly; SF-36v2 = Short Form 36 version 2 Health Survey version 2; SMBG = self-monitored blood glucose; VAS = Visual Analog Scale; VLDL-C = very low-density lipoprotein cholesterol.

Pharmacokinetic, pharmacodynamic, and PK/PD analyses are the responsibility of Lilly's PK/PD group, and therefore the detailed analysis plan will not be included in this SAP. A summary of tirzepatide concentration by time will be reported. Exposure-response analysis between tirzepatide concentration and safety, pharmacology, and efficacy may be performed using population PK and population PK/PD nonlinear mixed-effects modeling techniques implemented on NONMEM software. Additionally, the impact of intrinsic and extrinsic factors (such as age, weight, sex, renal, hepatic functions, and baseline histological scores) on PK and/or PD parameters may be evaluated. If ADA titers are detected in the immunogenicity samples, the impact of titers on tirzepatide PK and/or PD may be evaluated.

4.5. Exploratory Endpoints Analysis

Table GPHL.4.4. Exploratory Efficacy Analyses

Objective	Relative to the efficacy measure:	Analysis Conducted
compare tirzepatide 10 mg and/or 15 mg with placebo at 72 weeks	Percentage of participants achieving ≥25% body weight reduction from randomization	logistic model in Section 4.3.2.2
compare tirzepatide 10 mg and/or 15 mg with placebo at 72 weeks	Percentage of participants whose BMI shifts between clinically relevant categories, ie, from baseline (<25, 25 to <30, 30 to <35, 35 to <40, ≥40) to postbaseline (<25, 25 to <30, 30 to <35, 35 to <40, ≥40)	Shift analysis will be conducted based on data from EAS.
compare tirzepatide 10 mg and/or 15 mg with placebo at 72 weeks	Mean change in health outcome measurements (Section 4.7.1)	Terms of treatment, stratification factors, and baseline measurement score will be used as a covariate. Missing data will be imputed using LOCF method.

Abbreviations: BMI = body mass index; EAS = Efficacy Analysis Set; LOCF = last observation carried forward.

In addition, for the following endpoints, analyses to assess the risk difference in proportions for an unconditional treatment effect (Ge et al. 2011) among each tirzepatide dose and placebo arm may be conducted:

- percentage of participants achieving at least 5% body weight reduction at 72 weeks
- percentage of participants achieving at least 10% body weight reduction at 72 weeks
- percentage of participants achieving at least 15% body weight reduction at 72 weeks
- percentage of participants achieving at least 20% body weight reduction at 72 weeks
- percentage of participants achieving at least 25% body weight reduction at 72 weeks

4.6. Safety Analyses

Unless specified otherwise, safety assessments will be based on the SS (Table GPHL.3.1). All events that occur between the first dose date of study drug 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.

For selected continuous safety parameters, the mean change from baseline versus placebo will be assessed via a MMRM using REML. Data from scheduled visits will be utilized for this analysis. Unless specified otherwise, the model will include country/pooled country, sex, type of AHM used at randomization, treatment group, visit and treatment-by-visit interaction 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 4.3.2.2 will be tested in order until convergence is met. 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.

4.6.1. 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.

4.6.2. Adverse Events

4.6.2.1. Treatment-Emergent Adverse Events

A TEAE is defined as an event that first occurred or worsened in severity after baseline. 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 (eg, treatment emergent flag, start time of study treatment and event) will be used to determine whether the event was pre versus posttreatment, if available. If the relevant information is not available, then the events will be counted as posttreatment.

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. The SOC will be 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, SAE, death, discontinued from study treatment or study due to an AE, relationship to study drug 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 nonmissing severities. If all severities are missing for the defined postbaseline period of interest, it will show as missing in the table.

4.6.2.2. 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.

4.6.3. Additional Safety Assessment

4.6.3.1. Deaths

A listing of all deaths during the study will be provided. The listing will include participant identification including the treatment group, site number, participant number, date of death, age at the time of enrollment, sex, cause of death as reported by investigator, cause of death as adjudicated by CEC, etc.

4.6.3.2. 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 and participant number, date of event, age at the time of enrollment, sex, MedDRA SOC and PT, reported term, severity, outcome, relationship to study drug, time from first dose of study drug to the event, AE start date, AE end date, severity, and action taken related to study treatment.

4.6.3.3. Other Significant Adverse Events

The counts and percentages of participants who discontinued from study treatment 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.

4.6.3.4. Vital Signs

In the case that multiple records of an individual vital sign are collected at the same visit, they will be averaged prior to being used for data summaries and analyses.

Vital signs 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 4.6. Only scheduled measurements will be included in the mean change analyses.

Counts and percentages of participants with treatment-emergent abnormal (ie, high or low) sitting SBP, sitting DBP, and pulse at any time during the entire study (including the safety follow up period) will be presented 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 stated in Table GPHL.4.5.

Table GPHL.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.

In addition, the following analyses will be conducted by treatment:

- counts and percentages of participants who had resting heart rate increases from baseline at 2 or more consecutive visits of more than 10 bpm and/or 20 bpm
- counts and percentages of participants who had resting heart rate at any postbaseline visit increases from maximum baseline ≥20 bpm
- counts and percentages of participants who had at least 1 resting heart rate at any postbaseline visit exceeding 100 bpm, and
- counts and percentages of participants who had at least 1 resting heart rate exceeding 100 bpm occurring at 2 or more consecutive study visits.

4.6.3.5. Electrocardiograms

Summary statistics by treatment and by nominal visit will be provided for ECG parameters (heart rate, PR, QRS, QT, and QT corrected QTcF [QTcF = QT / $RR^{0.333}$]). When the QRS is prolonged (eg, a complete bundle branch block), QT and corrected QT interval should not 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 (heart rate and PR) will be summarized for participants who have both a baseline and at least 1 postbaseline result. Only scheduled measurements will be included in the mean change analyses.

The criteria for identifying participants with treatment-emergent quantitative ECG abnormalities is based on Table GPHL.4.6.

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

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

- treatment-emergent increase from the maximum baseline in QTcF interval of greater than 30 msec, 60 msec, or 75 msec. Maximum baseline will be the maximum nonmissing observation in the baseline period. The maximum value during the treatment period will be analyzed. Scheduled and unscheduled measurements will be included.

Table GPHL.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; QTcF = Fredericia's corrected QT interval.

4.6.3.6. Clinical Laboratory Evaluation

Limits from the performing lab will be used to define low and high. 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.

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 scheduled measurement. Baseline is defined in Table GPHL.4.1. Unscheduled 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.

For qualitative laboratory analytes, the number and percentage of participants with normal and abnormal values may be summarized by treatment.

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 or ANCOVA (if MMRM model is not applicable) will be used for the analysis during the treatment period for the continuous measurements for selected lab tests.

4.6.4. Patient Narratives

Patient narratives will be provided for all participants who experience any of the following "notable" events:

- death

- SAE
- permanent discontinuation of study treatment due to AEs
- pregnancy

Patient narratives (patient level data and descriptive paragraph) will be provided for participants in the randomized population with at least 1 notable event.

4.6.5. 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 participant profile. Adverse events of special interest are defined in each section of special safety topics, where applicable.

4.6.5.1. Exocrine Pancreas Safety

4.6.5.1.1. 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.

An MMRM analysis will be used to analyze each pancreatic enzyme with a log transformed (postbaseline measure/baseline measure) response variable and treatment, nominal visit, treatment-by-nominal visit interaction as fixed effects.

4.6.5.1.2. Pancreatitis Events

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

4.6.5.2. Gastrointestinal Safety

4.6.5.2.1. Nausea, Vomiting, and Diarrhea

Summaries and analyses for incidence and severity of nausea, vomiting (including “vomiting” and “vomiting projectile”), diarrhea (including “diarrhea” and “diarrhoea”), and 3 events combined, will be provided by each treatment group.

Summary of the prevalence over time for nausea, vomiting, and diarrhea will also be presented.

Time to the first onset of nausea, vomiting, and diarrhea will be plotted.

4.6.5.2.2. Severe Gastrointestinal Events

Severe GI AEs will be captured with the 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 locks will be used to identify GI AEs, and only the PTs with severe/serious 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.

4.6.5.3. Hepatobiliary Disorders

4.6.5.3.1. 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 6.6 (Section 6.6).

4.6.5.3.2. 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 6.6 (Section 6.6).

Severe/serious acute gallbladder disease will be considered as AESIs and summarized separately.

4.6.5.3.3. Liver Enzymes

Common analyses for laboratory analyte measurements described in Section 4.6.3.6 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 safety 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 nonmissing 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 will be summarized for all participants with a postbaseline value, and for subsets based on various levels of baseline value:

  - participants whose nonmissing 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 will be summarized for all participants with a postbaseline value and for subsets based on various levels of baseline value, as described above for TBL.

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

4.6.5.4. 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 patient 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. Plasma glucose 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/participant/year) of level 2 or level 3 hypoglycemia, and level 3 hypoglycemia will be reported by treatment group.

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. When the number of hypoglycemic events is less than 10, a listing of hypoglycemic events will be provided instead.

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 an SAE form.

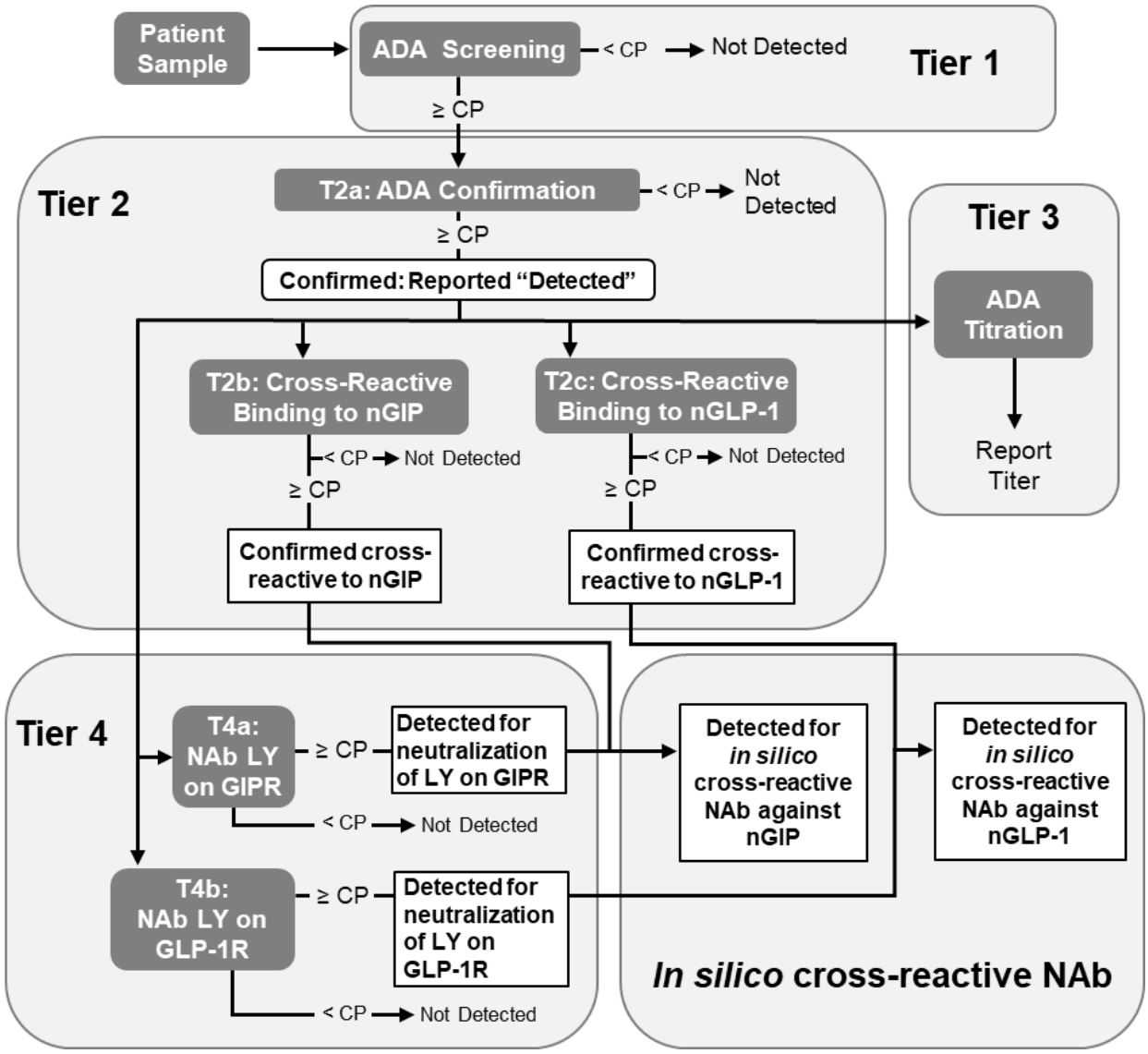
Severe/serious hypoglycemia will be considered as AESIs. The summaries of severe/serious hypoglycemia will be provided by treatment group. A listing of all events of severe/serious hypoglycemia may be provided, if deemed necessary. This listing will provide treatment allocation, clinical characteristics of the hypoglycemic event, and concomitant antihyperglycemic medications.

4.6.5.5. Immunogenicity

4.6.5.5.1. 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 GPHL.4.1 details a flow chart that reflects the multi-tiered testing approach.



Abbreviations: ADA = anti-drug antibodies; CP = cut point; GIPR = 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 GPHL.4.1. Flowchart of immunogenicity multitiered testing approach.

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

Table GPHL.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 GPHL.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 GPHL.4.8).

Table GPHL.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 (ie, 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 GPHL.4.1).

Similar terminology to Table GPHL.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 GPHL.4.9).

Table GPHL.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* Tier 4a: "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* Tier 4b: "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.

4.6.5.5.2. 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 nonmissing 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.

4.6.5.5.3. 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 GPHL.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.

4.6.5.5.4. 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 GPHL.4.10) by TE ADA status (TE ADA+, TE ADA-, TE ADA Inconclusive) during the planned treatment period.

Table GPHL.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 = Standardised 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 GPHL.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 injection site reactions 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.

4.6.5.6. 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 eCRF, only date (no time) information are collected; if such events occurred on the same date as the study drug injection date, they will be included in Time Period A.

Time Period B, of potential nonimmediate 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) (Note that Anaphylactic reaction SMQ algorithm will only be summarized for Time Period A)
- Narrow terms in Angioedema SMQ (20000024)
- Narrow terms in Severe cutaneous adverse reactions SMQ (20000020)
- Narrow terms in Hypersensitivity SMQ (20000214), and

- 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 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 (ie, combined search across narrow of all 5 SMQs)
- Any narrow term within each SMQ, separately (ie, 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.

4.6.5.6.1. 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.

4.6.5.7. 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 patient with a single statistic, typically an extreme value. This analysis allows each patient 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 patients. 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 patients to the analysis.

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

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

4.6.5.7.1. Severe/Serious Injection Site Reactions

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

The counts and percentage of participants with severe/serious treatment-emergent ISRs will be summarized by treatment. A listing of participants with treatment-emergent severe/serious ISRs may be provided, if deemed necessary.

4.6.5.8. 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.

4.6.5.9. 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 6.6 (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. A listing of participants with major depressive disorder/suicidal ideation or behavior may be provided if deemed necessary.

In addition to spontaneously reported AEs assessed by the investigator, suicidal ideation and behavior, and depression will be assessed through the use of the C-SSRS and the (PHQ-9), respectively.

4.6.5.9.1. Patient Health Questionnaire

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

- none (not depressed): 0 through 4
- mild: 5 through 9
- moderate: 10 through 14
- moderately severe: 15 through 19, and
- severe: 20 through 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 (that is, 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

4.6.5.9.2. 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), and
- 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 through 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 through 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 through 10) on the C-SSRS.

A listing containing data for each participant with suicidal ideation, suicidal behavior, or nonsuicidal self-injurious behavior during the study by treatment and visit 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.

4.6.5.10. Malignancy

Treatment-emergent malignancy will be considered 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 6.6 (Section 6.6).

4.6.5.11. Renal Safety

Laboratory measures related to renal safety will be analyzed as specified for laboratory measurements in Section 4.6.3.6.

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/1.73 m^2, using categories (<30, ≥30 to <45, ≥45 to <60, ≥60 to <90, 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.

4.6.5.11.1. 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) and
- 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.

4.6.5.11.2. Dehydration

Dehydration events will be captured in the Narrow terms in Dehydration SMQ (20000232). Severe/serious dehydration events will be considered AESIs.

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.

4.6.5.12. Thyroid Safety Monitoring

4.6.5.12.1. 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.

4.6.5.12.2. C-Cell Hyperplasia and Thyroid Malignancies

Treatment-emergent thyroid malignancies and C-Cell hyperplasia will be considered AESIs. Detailed searching criteria can be found in Appendix 6.6 (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.

4.6.5.13. Treatment-Emergent Arrhythmias and Cardiac Conduction Disorders

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

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

4.6.5.14. Overdose

Overdose is defined as taking more than 15 mg of tirzepatide in less than 72 hours. Overdosing of tirzepatide will be summarized by treatment group, and a listing of participants with tirzepatide overdosing will be provided.

In addition, a listing of participants reporting AEs related to overdosing of tirzepatide may be provided.

4.6.5.15. 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.

4.6.5.16. Amputation/Peripheral Revascularization

Treatment-emergent amputation/peripheral revascularization will be considered as AESIs. The counts and percentages of participants with amputations/peripheral revascularization may be summarized by treatment. Detailed searching criteria can be found in Appendix 6.6 (Section 6.6).

4.6.5.17. Diabetic Retinopathy Complications

Results of the baseline dilated fundoscopic exam will be summarized by treatment. Any TEAE suspected of worsening retinopathy triggers a follow-up dilated fundoscopic exam. A summary of TEAEs suspected of worsening retinopathy (Section 6.6) and a summary of the results of the follow-up dilated fundoscopic exam will be summarized by treatment and PT. The cases with repeated fundoscopy during the course of the trial, based on clinical suspicion of worsening retinopathy that have either findings of de novo retinopathy or progression of retinopathy, and severe/serious adverse events from the PTs defined in searching criteria in Appendix 6.6 (Section 6.6) will be considered as AESI and summarized.

4.6.5.18. Metabolic Acidosis

Severe/serious treatment-emergent metabolic acidosis, including diabetic ketoacidosis, will be captured as an AESI.

The counts and percentages of participants with metabolic acidosis, including diabetic ketoacidosis, may be summarized by treatment based on searching criteria in Appendix 6.6 (Section 6.6).

4.6.5.19. Persistent Hyperglycemia

A summary of initiation of rescue therapy in response to persistent hyperglycemia will be provided by treatment. If there are sufficient number of episodes (≥10), time-to-first-event analyses for the initiation of rescue therapy will be conducted by treatment using a cox proportional regression model. For patients without event "time-to-event," event time will be censored at end of treatment period. A listing of patients who initiated rescue therapy will be provided.

4.7. Other Analyses

4.7.1. 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.

4.7.1.1. Patient Global Impression of Status for Physical Activity

The counts and percentages of participants PGIS response categories 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.

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

Per copyright owner, the QualityMetric Health Outcomes™ Scoring (PRO_CoRe V2.0) Software 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 GPHL.4.3.

The empirical cumulative distribution function (eCDF) curves of the change from baseline to Week 72 in SF-36 physical function domain score will be provided by treatment group.

If data allowed, analysis for SF-36 physical function domain score analysis described in Table GPHL.4.3 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."

4.7.1.3. 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 follows:

- Each composite raw score will be calculated if a minimum of 50% of the items for that composite has a nonmissing 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 for a composite, then:
    - The average of the valid nonmissing 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 participant at each time point:

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

        - C_{avg} is the raw average score of all nonmissing 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 (that is, 5)
- S_{min} is the minimum possible raw score value (that is, 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 GPHL.4.4.

If data allows, analysis for IWQOL-Lite-CT physical function composite score analysis described in Table GPHL.4.4 will be conducted to evaluate the treatment effect in participants who have limitations in physical function at baseline (as defined in Section 4.7.1.2).

4.7.1.4. EQ-5D-5L

For the utility score and the Visual Analog Scale (VAS) scores, the following analyses of the actual value and change from baseline value will conducted:

- descriptive summaries by treatment group and
- ANCOVA analysis described in Table GPHL.4.4.

4.7.2. Subgroup Analyses

Efficacy subgroup analyses will be guided by the treatment-regimen estimand in FAS and the efficacy estimand in EAS.

Subgroup analyses may be done by country to support local regulatory registrations.

4.7.2.1. Subgroup Analysis of Body Weight Change

Subgroup analyses by the following baseline characteristics will be provided:

- age group (<65, ≥65 years)
- race
- sex
- ethnicity
- region of enrollment (US, outside of US)
- BMI group (<30, ≥30 and <35, ≥35 and <40, ≥40 kg/m^2), and
- baseline HbA1c group (≤8.5%, >8.5%)
- type of AHM used at randomization (weight loss, weight gain, weight neutral)

The outcome measures for the subgroup analyses will include:

- mean percent change in body weight from randomization at 72 weeks, and

- percentage of participants achieving at least 5% body weight reduction at 72 weeks.

For the percentage change in body weight from randomization at 72 weeks, for each subgroup analyses aforementioned, 2 models will be conducted as described below.

Treatment regimen estimand

- Conduct ANCOVA model on the subgroup only with terms of treatment group, country/pooled country, sex, and type of AHM used at randomization as fixed effects, and baseline body weight as a covariate. Missing body weight measurement at 72 weeks will be imputed using the imputation method described in Section 4.3.2.3. For the subgroup analysis on the type of AHM used at randomization, the variable type of AHM used at randomization will be removed from ANCOVA model as a fixed effect.
- Full ANCOVA model: treatment group, subgroup, treatment-by-subgroup-interaction, country/pooled country, sex, and type of AHM used at randomization as fixed effects, and baseline body weight as a covariate. Missing body weight measurement at 72 weeks will be imputed using the imputation method described in Section 4.3.2.3. For the subgroup analysis on the type of AHM used at randomization, the variable type of AHM used at randomization will be removed from ANCOVA model as a fixed effect.

Efficacy estimand

- Conduct MMRM model on the subgroup only with terms of treatment group, visit, treatment-by-visit-interaction, country/pooled country, sex, and type of AHM used at randomization 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. For the subgroup analysis on the type of AHM used at randomization, the variable type of AHM used at randomization will be removed from MMRM model as a fixed effect.
- Full MMRM model: treatment group, visit, subgroup, treatment-by-visit-interaction, treatment-by-subgroup-interaction, subgroup-by-visit-interaction, treatment-visit-subgroup-interaction, country/pooled country, sex, and type of AHM used at randomization 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. For the subgroup analysis on the type of AHM used at randomization, the variable type of AHM used at randomization will be removed from MMRM model as a fixed effect.

For the percentage of participants achieving at least 5% body weight reduction at 72 weeks, for each subgroup analyses aforementioned, 2 models will be conducted as described below.

Treatment regimen estimand

- Conduct logistic regression model on the subgroup with terms of treatment group, country/pooled country, sex, and type of AHM used at randomization as fixed effects, and baseline body weight as covariate. Missing body weight measurement at 72 weeks will be imputed using the imputation method described in Section 4.3.2.3, then the continuous measurements will be dichotomized into status of achieving at least 5% body weight reduction (Yes or No). For the subgroup analysis on the type of AHM used at randomization, variable type of AHM used at randomization will be removed from logistic regression model as a fixed effect.

- Conduct logistic regression model with terms of treatment group, subgroup, treatment-by-subgroup-interaction, country/pooled country, sex, and type of AHM used at randomization as fixed effects, and baseline body weight as a covariate. Missing body weight measurement at 72 weeks will be imputed using the imputation method described in Section 4.3.2.3, then the continuous measurements will be dichotomized into status of achieving at least 5% body weight reduction (Yes or No). For the subgroup analysis on the type of AHM used at randomization, variable type of AHM used at randomization will be removed from logistic regression model as a fixed effect.

Efficacy estimand

- Conduct logistic regression model on the subgroup with terms of treatment group, country/pooled country, sex, and type of AHM used at randomization as fixed effects, and baseline body weight as covariate. Missing body weight measurement at 72 weeks will be imputed by the predicted value from MMRM model on the subgroup aforementioned, then the continuous measurements will be dichotomized into status of achieving at least 5% body weight reduction (Yes or No). For the subgroup analysis on the type of AHM used at randomization, variable type of AHM used at randomization will be removed from logistic regression model as a fixed effect.
- Conduct logistic regression model with terms of treatment group, subgroup, treatment-by-subgroup-interaction, country/pooled country, sex, and type of AHM used at randomization as fixed effects, and baseline body weight as a covariate. Missing body weight measurement at 72 weeks will be imputed by the predicted value from full MMRM model aforementioned, then the continuous measurements will be dichotomized into status of achieving at least 5% body weight reduction (Yes or No). For the subgroup analysis on the type of AHM used at randomization, variable type of AHM used at randomization will be removed from logistic regression model as a fixed effect.

4.8. Interim Analyses

Not applicable.

4.9. Changes to Protocol-Planned Analyses

To provide data on efficacy of the investigational product that would be valuable to better inform clinical decisions in management of people living with obesity and T2DM, protocol-planned objectives are changed as below.

Key secondary objectives

- Added the key secondary objectives of "Demonstrate that tirzepatide 10 mg and/or 15 mg QW are superior to placebo for the percentage of participants who achieve ≥20% body weight reduction from randomization at 72 weeks".
- Changed the objectives "Demonstrate that tirzepatide 10 mg and/or 15 mg QW are superior to placebo for percentage of participants who achieve the target value of HbA1c ≤6.5% and <5.7% at 72 weeks" from additional secondary objectives to key secondary objectives.

- Added the key secondary objectives of "Demonstrate tirzepatide QW (all doses combined) is superior to placebo for change in HDL-cholesterol (mg/dL) and non-HDL-cholesterol (mg/dL)."
- Changed the objectives of "Demonstrate tirzepatide QW (all doses combined) is superior to placebo for change in total cholesterol (mg/dL), low-density lipoprotein cholesterol (LDL-cholesterol) (mg/dL)" from the key secondary objectives to additional secondary objectives.

Additional secondary objectives

- Added the objective of "Demonstrate tirzepatide 10 mg and/or 15 mg QW are superior to placebo in change from randomization for 7-point self-monitored blood glucose profiles at 72 weeks" as an additional secondary objective.
- Added the objective of "Demonstrate tirzepatide 10 mg and/or 15 mg QW are superior to placebo in change from randomization for EQ-5D-5L utility score and VAS score at 72 weeks" as an additional secondary objective.

5. Sample Size Determination

Approximately 1300 participants will be screened to achieve 900 randomly assigned to study intervention (300 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 to placebo, a common SD of 10%, and a dropout rate of 25% are assumed for statistical power calculations. Under the assumptions above, randomizing 900 participants in a 1:1:1 ratio to tirzepatide 10 mg (300), tirzepatide 15 mg (300), and placebo (300) 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 10-mg tirzepatide and 15-mg tirzepatide dose to placebo in terms of proportion of participants achieving at least 5% body weight reduction at 72 weeks, conducted in parallel using a Chi-square test, each at a 2-sided significance level of 0.025, assuming 25% placebo treated participants and 90% tirzepatide-treated participants achieving the goal and a dropout rate of 25%.

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics

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), race, ethnicity, height (cm), weight (kg), BMI (kg/m^2), waist circumference (cm), age group (<65 years, ≥65 years), BMI group (<30, ≥30 and <35, ≥35 and <40, ≥40 kg/m^2), blood pressure (mmHg), country, weight-related comorbidities, HbA1c, fasting serum glucose, lipids, type of AHM, and T2DM duration.

6.1.2. Historical Illnesses and Preexisting Conditions

The count and percentages of participants with historical illnesses and preexisting conditions will be summarized by treatment group using the MedDRA PTs nested within SOC. The SOC will be in alphabetical order. Conditions (ie, PTs) will be ordered by decreasing frequency within SOC. This will be summarized for all randomized participants.

6.1.3. Concomitant Therapy

Concomitant medication will be summarized by PTs by treatment group by decreasing frequency for SS group.

Additionally, medications of interest (as defined below) will be summarized by treatment for SS.

Concomitant medications of interest include the following:

- baseline antihypertensive therapy
- baseline lipid lowering therapy
- baseline antihyperglycemic therapy
- use of other antihyperglycemic therapy after randomization
- rescue therapy due to persistent hyperglycemia
- changes to baseline medication post-randomization:
    - antihypertensive therapy
    - lipid lowering therapy, and
    - antihyperglycemic therapy.
- utilization after randomization of:
    - antidiarrheal medication, and
    - antiemetic medication.

6.2. Appendix 2: Treatment Compliance

Summary of premature discontinuation of study treatment (including discontinuation reason) will be provided by study treatment. A time-to-event analysis of premature study treatment discontinuation will also be conducted.

If data warrant, the counts and percentages of participants who follow the planned escalation scheme, have dose interruption, or have dose de-escalation may be summarized for tirzepatide treatment groups. 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 randomized tirzepatide treatment and visit during the dose escalation period.

Treatment compliance will be defined as taking at least 75% of the scheduled tirzepatide doses. Compliance over the study 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 study period. Treatment compliance will be summarized descriptively in the study period by treatment using the mITT population.

6.3. Appendix 3: Important Protocol Deviations

Important protocol deviations are defined in the Trial Issues Management Plan. 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 requirements.

Analyses provided for the Clinical Trial Registry requirements include the following:

- Summary tables of AEs, provided as datasets which will be converted to an XML files. S will be summarized by treatment group and MedDRA PT.
- 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, and
    - the number of events experienced.
- For each SAE, these additional terms are provided for EudraCT:
    - the total number of occurrences causally related to treatment
    - the total number of deaths, and
    - the total number of deaths causally related to treatment.

Demographic table including the following age ranges required by EudraCT: 18 to 65 years, 65 to 85 years, and 85 years and over.

6.5. Appendix 5: COVID-19 Pandemic Impact

This section lists additional statistical analyses that may be performed at final 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 6.6 (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 in Study GPHL, 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 and key secondary endpoints, missing data due to exceptional circumstances will be handled as described in Section 4.3.2.3. In addition, a summary table for participants whose primary or key secondary measurements were impacted by COVID-19 (including missing, collected using alternative options) may be provided. A listing of participants whose primary or key secondary 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.

Diabetic Retinopathy Complications (preferred terms)

Arteriosclerotic retinopathy, Blindness, Choroidal neovascularisation, Cystoid macular oedema, Detachment of macular retinal pigment epithelium, Detachment of retinal pigment epithelium, Diabetic blindness, Diabetic eye disease, Diabetic retinal oedema, Diabetic retinopathy, Diabetic uveitis, Exudative retinopathy, Eye laser surgery, Fundoscopy, Fundoscopy abnormal, Intra-ocular injection, Macular detachment, Macular oedema, Maculopathy, Noninfective chorioretinitis, Noninfective retinitis, Phacotrabeculectomy, Retinal aneurysm, Retinal arteriovenous malformation, Retinal artery embolism, Retinal artery occlusion, Retinal artery stenosis, Retinal collateral vessels, Retinal cryoablation, Retinal detachment, Retinal exudates, Retinal aemorrhage, Retinal laser coagulation, Retinal neovascularisation, Retinal oedema, Retinal operation, Retinal thickening, Retinal vascular disorder, Retinal vascular occlusion, Retinal vein occlusion, Retinitis, Retinopathy, Retinopathy haemorrhagic, Retinopathy hypertensive, Retinopathy hyperviscosity, Retinopathy proliferative, Venous stasis retinopathy, Vitrectomy, Scintillating scotoma, Vision blurred, Visual impairment, Blindness transient, Blindness unilateral, Sudden visual loss, Visual acuity reduced, Visual acuity reduced transiently, Diplopia, Amaurosis, Amaurosis fugax

Note: This table was developed for MedDRA version 25.0 and will be updated to the version of MedDRA to be used at the time of the final database lock.

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 4 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” MedDRA SMQ (20000022, narrow scope) and a “Chronic pancreatitis” PT search of the AE database, while adjudication-confirmed pancreatitis are found from adjudication forms.

Malignancy

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.

Major Depressive Disorder/Suicidal Ideation

AEs will be searched using MedDRA PTs from SMQs narrow scope: 20000037 (Suicide/self-injury) and 20000167 (Depression [excluding suicide and self injury]).

Metabolic Acidosis, Including Diabetic Ketoacidosis

Metabolic acidosis including diabetic ketoacidosis will be identified using the following MedDRA 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.

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:

- 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.

- Cardiac Conduction Disorders
  - Conduction defects SMQ (20000056), narrow terms only; and
  - Cardiac conduction disorders High Level Term (HLT; 10000032), all PTs.

7. References

[ADA] American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes-2020. *Diabetes Care.* 2020;43(suppl 1):S66–S76. https://doi.org/10.2337/dc20-S006

[ICH] International Council for Harmonisation. Addendum on estimands and sensitivity analysis in clinical trials. To the guideline principles for clinical trials E9(R1): guidance for industry. November 2019. Accessed August 31, 2022. https://www.gmp-compliance.org/files/guidemgr/E9-R1_Step4_Guideline_2019_1203.pdf

Ge M, Durham LK, DR Meyer, et al. Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. *Drug Inf J.* 2011;45(4):481-493. https://doi.org/10.1177/009286151104500409

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

Qu Y, Dai B. Return-to-baseline multiple imputation for missing values in clinical trials. *Pharm Stat*. 2022;21(3):641-653. https://doi.org/10.1002/pst.2191

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

Signature Page for VV-CLIN-076568 v1.0

Approval	PPD 13-Jan-2023 12:47:41 GMT+0000

Signature Page for VV-CLIN-076568 v1.0