I8F-MC-GPHM Statistical Analysis Plan Version 3

Efficacy and Safety of Tirzepatide Once Weekly Versus Placebo After an Intensive Lifestyle Program in Participants Without Type 2 Diabetes Who Have Obesity or Are Overweight With Weight-Related Comorbidities: A Randomized, Double Blind, Placebo-Controlled Trial (SURMOUNT-3)

NCT04657016

Approval Date: 03-Feb-2023

Title Page

Protocol Title: Efficacy and Safety of Tirzepatide Once Weekly versus Placebo After an Intensive Lifestyle Program in Participants without Type 2 Diabetes who have Obesity or are Overweight with Weight-Related Comorbidities: A Randomized, Double-Blind, Placebo-Controlled Trial (SURMOUNT-3)

Protocol Number: I8F-MC-GPHM

Compound Number: LY3298176

Short Title: Effect of Tirzepatide versus Placebo after Intensive Lifestyle Program (SURMOUNT-3)

Acronym: SURMOUNT-3

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

IND 139721

Document ID: VV-CLIN-076448

Table of Contents

Title Page **1**

Table of Contents **2**

List of Abbreviations **6**

Version History **9**

1. Introduction **10**
1.1. Objectives, Endpoints, and Estimands 11
1.1.1. Objectives and Endpoints 11
1.1.2. Estimands 13
1.2. Study Design 16

2. Statistical Hypotheses **18**
2.1. Multiplicity Adjustment 18

3. Analysis Sets **20**

4. Statistical Analyses **21**
4.1. General Considerations 21
4.2. Participant Dispositions 22
4.3. Primary Endpoints Analysis 23
4.3.1. Definition of Endpoint 23
4.3.2. Main Analytical Approach 23
4.4. Secondary Endpoint(s) Analysis 26
4.4.1. Key Secondary Analyses Subject to Type 1 Error Rate Control 26
4.4.2. Supportive Secondary Endpoints 27
4.5. Exploratory Endpoints Analysis 30
4.6. Safety Analyses 32
4.6.1. Extent of Exposure 33
4.6.2. Adverse Events 34
4.6.3. Additional Safety Assessments 35
4.6.4. Patient Narratives 38
4.6.5. Special Safety Topics 38
4.7. Other Analyses 54
4.7.1. Health Outcomes 54
4.7.2. Subgroup Analyses 56
4.8. Interim Analyses 58
4.9. Changes to Protocol-Planned Analyses 58

5. Sample Size Determination **59**

6. Supporting Documentation **60**
6.1. Appendix 1: Demographic and Baseline Characteristics 60
6.1.1. Patient Characteristics 60
6.1.2. Historical Illnesses and Preexisting Conditions 60
6.1.3. Concomitant Therapy 60
6.2. Appendix 2: Treatment Compliance 61
6.3. Appendix 3: Important Protocol Deviations 61

Approved on 03 Feb 2023 GMT

6.4. Appendix 4: Analyses for the Intensive Lifestyle Modification Lead-in Period....61
6.4.1. General Consideration....61
6.4.2. Patient Characteristics....62
6.4.3. Disposition....62
6.4.4. Medical History, Preexisting Conditions, and Concomitant Medications....62
6.4.5. Adverse Events....62
6.4.6. Additional Safety Assessments....63
6.4.7. Important Protocol Deviations....63
6.5. Appendix 5: Clinical Trial Registry Analyses....63
6.6. Appendix 6: Exceptional Circumstances Impact....63
6.6.1. General Considerations....64
6.6.2. Exposure....64
6.6.3. Protocol Deviations....64
6.6.4. Patient Disposition....64
6.6.5. Adverse Events....64
6.6.6. Major Depressive Disorder/Suicidal Ideation....64
6.6.7. Local Laboratory Measurements....65
6.6.8. Missing Data Due to Exceptional Circumstances....65
6.7. Appendix 7: Searching Criteria for Special Safety Topics....65
7. References....67

Table of Contents

Table		Page
Table GPHM.1.1.	Objectives and Endpoints	11
Table GPHM.1.2.	Attribute Description of Estimands for Secondary Endpoints	15
Table GPHM.3.1.	Description of Analysis Datasets	20
Table GPHM.4.1.	Baseline and Postbaseline Definitions for Safety Groups	22
Table GPHM.4.2.	Secondary Measures Controlled for Type 1 Error	26
Table GPHM.4.3.	Secondary Measures Not Controlled for Type 1 Error	27
Table GPHM.4.4.	Exploratory Efficacy Analysis	30
Table GPHM.4.5.	Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurements	36
Table GPHM.4.6.	Selected Categorical Limits for ECG Data	37
Table GPHM.4.7.	Sample Anti-Drug Antibodies (ADA) Assay Results	43
Table GPHM.4.8.	Sample Clinical Anti-Drug Antibodies Interpretation Results	43
Table GPHM.4.9.	*In Silico* Classification for Cross-Reactive NAb	44
Table GPHM.4.10.	Adverse Events for Analysis with Immunogenicity Results	47
Table GPHM.6.1.	Baseline and Postbaseline Definitions in Lead-in Period	61

Approved on 03 Feb 2023 GMT

Table of Contents

Figure **Page**

Figure GPHM.1.1. Illustration of study design for Clinical Protocol I8F-MC-GPHM. 16
Figure GPHM.2.1. Type 1 error control strategy for primary and key secondary efficacy endpoints. 19
Figure GPHM.4.1. Flowchart of immunogenicity multitiered testing approach. 42

List of Abbreviations

Term	Definition
ADA	anti-drug antibody
AE	adverse event
AESI	adverse event(s) of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
BG	blood glucose
BMI	body mass index
BP	Bodily Pain
CEC	Clinical Evaluation Committee
CI	confidence interval
CRF	case report form
CSR	clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
CTR	Clinical Trial Registry
DBP	diastolic blood pressure
EAS	Efficacy Analysis Set
eCRF	electronic case report form
ED	early discontinuation
eGFR	estimated glomerular filtration rate
FAS	Full Analysis Set
GI	gastrointestinal
GIP	glucose-dependent insulinotropic polypeptide
GIPR	glucose-dependent insulinotropic polypeptide receptor
GLP-1	glucagon-like peptide-1
GLP-1R	glucagon-like peptide-1 receptor

Approved on 03 Feb 2023 GMT

H	Hypothesis
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
MACE	major adverse cardiovascular event(s)
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent-to-treat
MMRM	mixed model for repeated measures
MRD	minimum required dilution
MTD	maximum tolerated dose
NAb(+/-)	neutralizing antibody (positive/negative)
PF	Physical Function(ing)
PG	plasma glucose
PGIS	Patient Global Impression of Status for Physical Activity
PHQ-9	Patient Health Questionnaire
PK	pharmacokinetic(s)
PT	Preferred Term
QTcF	Fridericia's corrected QT interval
QW	once weekly
REML	restricted maximum likelihood
SAE	serious adverse event
SBP	systolic blood pressure
SD	standard deviation

Approved on 03 Feb 2023 GMT

SF-36v2	Short Form-36 Version 2 Health Survey acute form
SI	Système International
SMQ	Standardized MedDRA Query
SOC	System Organ Class
SS	Safety Analysis Set
TBL	total bilirubin
TE ADA (+/-)	treatment-emergent anti-drug antibody (positive/negative)
TEAE	treatment-emergent adverse event
TIMP	Trial Issues Management Plan
T2DM	type 2 diabetes mellitus
UACR	urine albumin-to-creatinine ratio
ULN	upper limit of normal

Version History

This Statistical Analysis Plan (SAP) for Study I8F-MC-GPHM (GPHM) is based on the protocol dated 19 November 2020.

SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	25 Mar 2021	Not applicable	Original version
2	07 Oct 2022	Details can be found in SAP Version History	Details can be found in SAP Version History
3	See date on Page 1	See below	See below

The key changes of the SAP Version 3 are summarized below:

1. In response to US Food and Drug Administration (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).
2. 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).
3. 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).
4. Updated to add body weight at Visit 10 (Week 0) as an additional covariate to the mixed model for repeated measures (MMRM) models for the analyses of mean change in body weight, mean percent change in body weight, and mean change in body mass index (BMI) from Week -12 to Week 72. This covariate adjusts for body weight prior to randomization on the final outcome (see Section 4.4.2.1).
5. Added exploratory endpoints to assess risk difference between tirzepatide maximum tolerated dose (MTD) and placebo arm for an unconditional effect in proportions of participants achieving body weight reduction targets (see Section 4.5).

1. Introduction

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

1.1. Objectives, Endpoints, and Estimands

1.1.1. Objectives and Endpoints

Table GPHM.1.1. Objectives and Endpoints

Objectives	Endpoints
Primary	
To demonstrate that tirzepatide MTD is superior to placebo from randomization for the following (measured at 72 weeks): • Percent change in body weight AND • Proportion of participants with ≥5% body weight reduction	• Mean percent change in body weight • Percentage of study participants who achieve ≥5% body weight reduction
Key Secondary (controlled for type I error)	
To demonstrate that tirzepatide MTD is superior to placebo from randomization for the following (measured at 72 weeks):	
• Maintaining body weight reduction achieved during the 12-week lead-in period	• Percentage of study participants who maintain ≥80% of the body weight lost during the 12-week lead-in period
• Body weight	• Percentage of study participants who achieve: o ≥10% body weight reduction o ≥15% body weight reduction o ≥20% body weight reduction
• Waist circumference	• Mean change in waist circumference (cm)
Additional Secondary	
Compare tirzepatide MTD QW with placebo from randomization for the following (measured at 72 weeks):	
• Body weight and BMI	• Mean change in body weight (kg) • Mean change in BMI (kg/m^2)
• Blood pressure	• Mean change in o systolic blood pressure (mmHg) o diastolic blood pressure (mmHg)

Approved on 03 Feb 2023 GMT

Objectives	Endpoints
• Lipid parameters	• Mean change in: o Total cholesterol (mg/dL) o HDL-C (mg/dL) o Non-HDL-C (mg/dL) o LDL-C (mg/dL) o VLDL-C (mg/dL) o Triglycerides (mg/dL) o Free fatty acids (mg/dL)
• Glycemic control	• Mean change in o fasting glucose (mg/dL) o HbA1c (%)
• 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
Compare tirzepatide MTD QW with placebo from Visit 2 for the following (measured at 72 weeks):	
• Body weight and BMI	• Mean change in absolute body weight (kg) • Mean percent change in body weight • Mean change in BMI (kg/m^2)
• Waist circumference	• Mean change in waist circumference (cm)

Abbreviations: BMI = body mass index; HbA1c = hemoglobin A1c; HDL-C = high-density lipoprotein cholesterol; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite-Clinical Trials Version; LDL-C = low-density lipoprotein cholesterol; MTD = maximum tolerated dose; QW = once weekly; SF-36v2 acute form = Short Form-36 Version 2 Health Survey acute form; VLDL-C = very low-density lipoprotein cholesterol.

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 the once weekly (QW) tirzepatide MTD (10 mg 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 without type 2 diabetes mellitus (T2DM), who have obesity (BMI ≥30 kg/m^2), or are overweight (BMI ≥27 kg/m^2) with at least 1 weight-related comorbid condition?

For Primary Disclosure of Study GPHM, Including Submission to the US FDA

The estimand is described by the following attributes:

- Population - Participants qualifying for study entry and randomization per protocol criteria.
- Co-primary endpoints - Percent change, from randomization to 72 weeks, in body weight AND participant status of achieving ≥5% body weight reduction from randomization to 72 weeks.
- Treatment condition of interest - Tirzepatide MTD versus placebo, regardless of study drug adherence.
- Handling of intercurrent events (ICEs) - The ICEs leading to treatment discontinuation for any reason are addressed by the treatment condition of interest attribute and handled by treatment policy strategy as described in the International Council for Harmonisation (ICH) E9 (R1) Addendum. 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

The estimand was requested by the FDA. It aims at reflecting how participants with obesity, or overweight with weight-related comorbidities, are treated in clinical practice, and it considers both safety and efficacy.

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

For <u>All Other Purposes</u>

The estimand is described by the following attributes:

- Population - Participants qualifying for study entry and randomization per protocol criteria.
- Co-primary endpoints - Percent change from randomization to 72 weeks in body weight AND participant status of achieving ≥5% body weight reduction from randomization to 72 weeks.

- Treatment condition of interest - Tirzepatide MTD versus placebo, excluding data after discontinuation of study drug.
- Handling of ICEs - The ICEs leading to treatment discontinuation for any reason are addressed by the treatment condition of interest attribute and handled by the hypothetical strategy as described in the ICH E9 (R1) Addendum.
- 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 ICEs leading to treatment discontinuation would not occur.

This estimand is referred to as the "efficacy" estimand in the latter sections.

1.1.2.2. Estimands for Secondary Endpoints

Table GPHM.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 GPHM.1.1	For submission to the US FDA, follow the same manner of the treatment regimen estimand in Section 1.1.2.1; for all other purposes, the same manner of the efficacy estimand in Section 1.1.2.1	For 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
Other secondary related to mean change	Same as estimands for co-primary endpoints	As described in Table GPHM.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

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

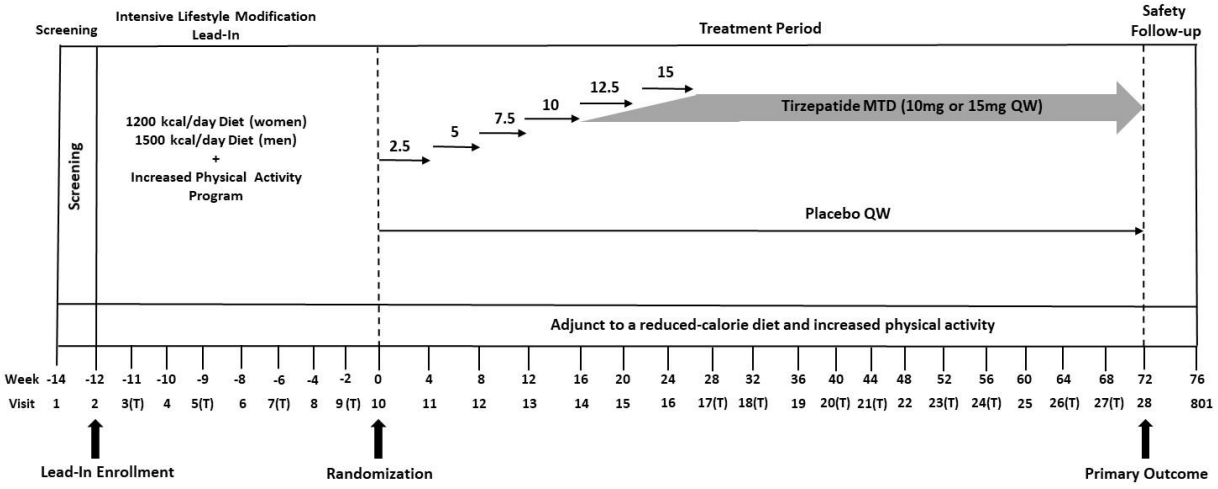
1.2. Study Design

Study GPHM is a Phase 3, multicenter, randomized, parallel-arm, placebo-controlled, double-blinded, 84-week study that will investigate the effects of treatment with the MTD of tirzepatide (10 mg or 15 mg QW) compared with placebo on body weight management of study participants who have obesity (BMI ≥30 kg/m^2), or are overweight (BMI ≥27 kg/m^2) with at least 1 weight-related comorbid condition, and achieved ≥5% weight reduction after a 12-week lead-in period with an intensive lifestyle modification program.

Study GPHM will consist of 4 periods:

- a 2-week screening period
- a 12-week lead-in period during which participants undertake an intensive lifestyle modification program to achieve ≥5% body weight reduction
- a 72-week, double-blind, placebo-controlled treatment period (including a 20-week dose escalation period), and
- a 4-week safety follow-up period.

Study participants who achieve at least 5% weight reduction will be randomized in a 1:1 ratio (tirzepatide MTD QW or placebo) at the end of the lead-in period. An upper limit of 70% enrollment of women will be used to ensure a sufficiently large sample of men.



Abbreviations: MTD = maximum tolerated dose; QW = once weekly; (T) = telephone visit.

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

The details about the overview of study periods and study visits can be found in the GPHM Protocol Section 4. The details of the unique visits not displayed in Figure GPHM.1.1 are provided below.

Visit 99

Visit 99 is only applicable to participants who discontinue the study treatment prematurely (between Visit 10 and Visit 28) and decline to complete the remaining scheduled study visits. These 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 endpoint.

Early Discontinuation of Treatment Visit

Participants unable or unwilling to continue the study intensive lifestyle intervention (during the lead-in period), or study drug intervention (during treatment period) for any reason, will perform an early discontinuation (ED) of treatment visit. If the participant is discontinuing during an unscheduled visit or a scheduled visit, that visit should be performed as the ED visit.

2. Statistical Hypotheses

The alternative hypothesis (H) for the primary objective is the following:

- H1 - QW tirzepatide MTD is superior to placebo for mean percent change in body weight AND the proportion of participants who achieve ≥5% body weight reduction from randomization to 72 weeks.
    - The above hypothesis will be tested at 2-sided significance level of 0.05.

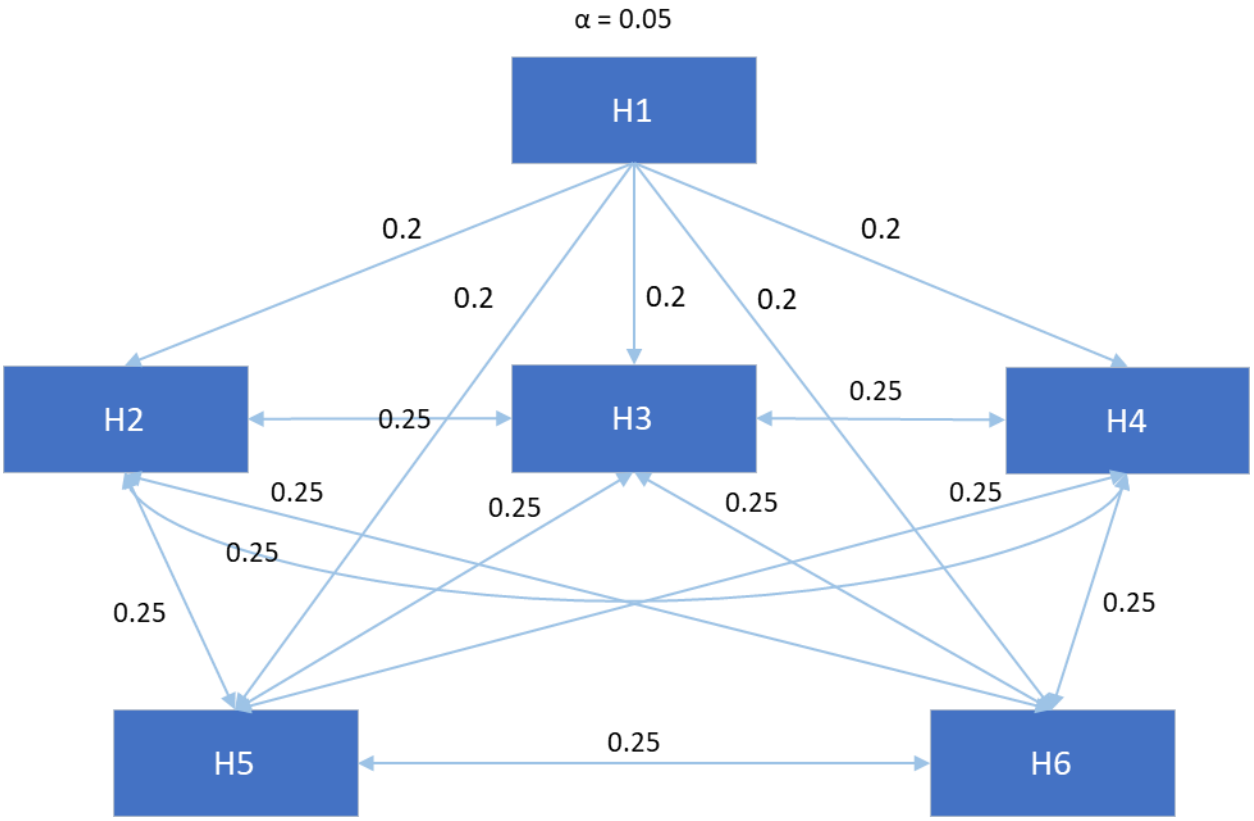
The alternative hypotheses for the key secondary objectives controlling for type 1 error rate are the following:

- H2 - QW tirzepatide MTD is superior to placebo for the proportion of participants who maintain ≥80% of the body weight lost during the 12-week lead-in period (per evaluation at 72 weeks).
- H3 - QW tirzepatide MTD is superior to placebo for the proportion of participants who achieve ≥10% body weight reduction from randomization to 72 weeks.
- H4 - QW tirzepatide MTD is superior to placebo for the proportion of participants who achieve ≥15% body weight reduction from randomization to 72 weeks.
- H5 - QW tirzepatide MTD is superior to placebo for the proportion of participants who achieve ≥20% body weight reduction from randomization to 72 weeks.
- H6 - QW tirzepatide MTD is superior to placebo for the mean change in waist circumference (cm) from randomization to 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 GPHM.2.1. Both efficacy estimand and treatment regimen estimand will be used to assess those 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 and safety assessments.

Approved on 03 Feb 2023 GMT



Abbreviation: H = hypothesis.

Figure GPHM.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 GPHM.3.1. Description of Analysis Datasets

Population	Description
Entered	All participants who sign informed consent
Enrolled	All participants who are enrolled in the 12-week lead-in period
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)	Data obtained during the double-blind treatment period from the mITT population, excluding data after discontinuation of study drug (last dose date + 7 days)
Full Analysis Set (FAS)	Data obtained during the treatment period from the mITT population, regardless of adherence to study drug
Safety Analysis Set (SS)	Data obtained during the treatment plus follow-up period from the mITT population, regardless of adherence to study drug

Note: 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, and 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 Eli Lilly and Company (Lilly) or its designee. Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (for example, few events to justify conducting an analysis). Additional analyses of the data may be conducted as deemed appropriate.

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

Unless specified otherwise, efficacy and safety will be assessed using the modified intent-to-treat (mITT) population, and data will be analyzed based on the assigned treatment (that is, not the actual treatment received by the participant). For submission to the US FDA to support the registration of tirzepatide for chronic weight management, the primary efficacy analysis will be conducted using the Full Analysis Set (FAS). For other purposes, the efficacy analysis of the double-blind period will be conducted using the Efficacy Analysis Set (EAS). Safety analysis of the double-blind period will be conducted using the Safety Analysis Set (SS).

Unless otherwise specified, for analyses in the double-blind period, baseline is defined as the last nonmissing data collected prior to or at randomization (prior to the first dosing of study drug). For analyses in the lead-in period, baseline is defined as the start of screening and ends prior to the intensive lifestyle modification program (typically at Week -12).

Summary statistics for continuous measures will include sample size, mean, standard deviation (SD), median, minimum, and maximum. The analysis model to make comparisons among treatment groups relative to continuous measurements assessed over time will be either analysis of covariance (ANCOVA) or an MMRM, with terms of treatment, visit, treatment-by-visit interaction, stratification factors, and corresponding baseline measurement as a covariate. An unstructured covariance structure will model the relationship of within-patient errors.

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

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

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

For some specific safety-related parameters, the definition of baseline and postbaseline for the double-blind period is specified in Table GPHM.4.1.

The definition of baseline and postbaseline during the intensive lifestyle modification lead-in period can be found in Appendix 4 (Section 6.4).

Table GPHM.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 ongoing at the time of the first dose of study drug in the double-blind treatment period.	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	The last nonmissing assessment (scheduled or unscheduled) recorded at or prior to randomization to the double-blind treatment period	Postbaseline will be defined as measurements after the randomization visit. 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 nonmissing assessment (scheduled or unscheduled) recorded during the 12-week lead-in period	Postbaseline will be defined as above (1.2). Only scheduled visits will be included. The early discontinuation visits are considered scheduled visits.

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

[a] Immunogenicity-related analysis is specified in Section 4.6.5.6.

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 the schedule of activity, where the last-observation-carried-forward (LOCF) approach will be applied to impute the endpoint when ED measure is available.

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

Statistical summaries and results of statistical analyses will be displayed in the following order: Placebo, tirzepatide MTD.

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

4.2. Participant Dispositions

A detailed description of participant disposition will be provided at the end of the study.

Frequency counts and percentages of all participants screened, enrolled, randomized, and receiving at least 1 dose of study drug will be summarized by treatment group, if applicable. Of the randomized population, frequency counts and percentages of participants who completed the study or prematurely discontinued the study (and/or study drug), including the reason for premature discontinuation, will be summarized by treatment group.

The study completion status for the randomized population is defined as following: participants with status reported via electronic case report form (eCRF) as "Completed" at the follow-up visit (Visit 801) will be considered as completers; otherwise, participants will be considered as noncompleters. A Kaplan-Meier analysis of time from randomization to premature discontinuation from the study and/or study treatment by treatment group will be provided.

A similar summary for the intensive lifestyle modification lead-in period can be found in Appendix 4 (Section 6.4).

4.3. Primary Endpoints Analysis

For submission of Study GPHM 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 multiple 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

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

$$(postbaseline\ body\ weight\ [kg] - baseline\ body\ weight\ [kg]) \\ / baseline\ body\ weight\ [kg] * 100.$$

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

4.3.2. Main Analytical Approach

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, and sex 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). For the

logistic regression, missing body weight data at 72 weeks will be imputed first, based on 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, the 2-sided 95% CI for mean change in percentage of body weight from baseline to the Week 72 visit between tirzepatide MTD and placebo will be derived.

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

4.3.2.2. The Analysis Related to the Efficacy Estimand

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

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

For the percentage of participants achieving at least 5% body weight reduction from randomization to the 72 weeks visit, a logistic regression model will be used with the response variable of the percentage of participants achieving at least 5% body weight reduction (Yes or No), with the missing value imputed using the predicted value from MMRM analysis followed by dichotomization. The terms for the logistic regression model will be the same as those for the treatment regimen estimand.

For the MMRM model, the independent variables of analysis model are treatment group, visit, and treatment-by-visit interaction; stratification factors (country and sex) as fixed effects; and baseline body weight as a covariate. An unstructured covariance structure will model the 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, the 2-sided 95% CI for mean percent change in body weight from randomization to the 72 weeks visit for tirzepatide MTD compared to placebo will be derived and summarized. The resulting least-squares-mean 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, the 2-sided 95% CI and odds ratio for percentage of participants achieving at least 5% body weight reduction from randomization to the 72-week visit for tirzepatide MTD compared to placebo will be summarized.

4.3.2.3. Methods for Hybrid Imputations

For efficacy analyses relative to the treatment regimen estimand, the ICEs and the resulting missing values 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), missing data are considered as missing at random. The missing data will be imputed using all non-missing 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 ED 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 statistical analysis software [SAS] program does not converge), an alternative multiple imputation method referencing 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 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 GPHM to the US FDA tirzepatide application for chronic weight management, additional sensitivity analyses of the primary efficacy outcomes will be conducted using the FAS and guided by 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 randomization 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 and Dai 2022).

4.4. Secondary Endpoint(s) Analysis

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

Table GPHM.4.2. Secondary Measures Controlled for Type 1 Error

Objectives	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to	Additional Information
Tirzepatide MTD QW is superior to placebo from randomization for maintaining body weight reduction achieved during the 12-week lead-in period (measured at 72 weeks)	Percentage of study participants who maintain ≥80% of the body weight lost during the 12-week lead-in period	Logistic model in Section 4.3.2.1 for treatment regimen estimand and logistic model in Section 4.3.2.2 for efficacy estimand	
Tirzepatide MTD QW is superior to placebo from randomization for body weight reduction (measured at 72 weeks)	Percentage of participants achieving body weight reduction ≥10%, ≥15%, and ≥20% from randomization 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	
Tirzepatide MTD QW is superior to placebo from randomization for waist circumference reduction (measured at 72 weeks)	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	LSM estimates will be plotted by treatment through 72-weeks.

Abbreviations: ANCOVA = analysis of covariance; LSM = least squares mean; MMRM = mixed model for repeated measures; MTD = maximum tolerated dose; QW = once weekly.

Decision will be guided by the 2-sided p-values in each objective, and details will be included in Section 2.1.

For submission of Study GPHM 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 the primary analysis. Assessment of the key secondary objective 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 the primary analysis.

4.4.2. Supportive Secondary Endpoints

Unless otherwise specified, other secondary and exploratory efficacy analyses will be guided by the efficacy estimand, using the same population as the primary analysis.

Analyses for labs will be performed for both Système International (SI) units and conventional (CN) units.

4.4.2.1. Additional Secondary Efficacy Analyses

Table GPHM.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 MTD QW with placebo from randomization for body weight and BMI reduction (measured 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	LSM estimates through 72 weeks will be plotted by study treatment.
Compare tirzepatide MTD QW with placebo from randomization for blood pressure reduction (measured at 72 weeks)	Mean change in systolic blood pressure (mmHg)	MMRM model in Section 4.3.2.2	LSM estimates through 72 weeks will be plotted by study treatment.
	Mean change in diastolic blood pressure (mmHg)	MMRM model in Section 4.3.2.2	LSM estimates through 72 weeks will be plotted by study treatment.
Compare tirzepatide MTD QW with placebo from randomization for improvement in lipid parameters (measured at 72 weeks)	Mean change in total cholesterol, HDL-C, non-HDL-C, LDL-C, VLDL-C, triglycerides, and free fatty acids from randomization	MMRM model in Section 4.3.2.2	Estimated means through 72 weeks will be plotted by study treatment. Log transformation will be adopted for lipid parameters.

Objective	Relative to the Efficacy Measure	Analysis Conducted in a Manner Similar to	Additional Information
Compare tirzepatide MTD QW with placebo from randomization for improvement of glycemic control (measured at 72 weeks)	Mean change in fasting glucose from randomization	MMRM model in Section 4.3.2.2	LSM estimates through 72 weeks will be plotted by study treatment.
	Mean change in HbA1c from randomization	MMRM model in Section 4.3.2.2	LSM estimates through 72 weeks will be plotted by study treatment.
Compare tirzepatide MTD QW with placebo from randomization for insulin reduction (measured at 72 weeks)	Mean change in fasting insulin	MMRM model in Section 4.3.2.2	Estimated means through 72 weeks will be plotted by study treatment. Log transformation will be adopted for fasting insulin.
Compare tirzepatide MTD QW with placebo from randomization for improvement of patient-reported outcomes (measured at 72 weeks)	Mean change in SF-36v2 acute form Physical Functioning (PF) domain score from randomization	ANCOVA model	Use terms of treatment, stratification factors, and baseline SF-36v2 PF score as a covariate. Missing data will be imputed using LOCF.
Compare tirzepatide MTD QW with placebo from randomization for improvement of patient-reported outcomes (measured at 72 weeks)	Mean change in IWQOL-Lite-CT PF composite score from randomization	ANCOVA model	Use terms of treatment, stratification factors, and baseline IWQOL-Lite-CT PF score as a covariate. Missing data will be imputed using LOCF.
Compare tirzepatide MTD QW with placebo from Visit 2 for body weight and BMI reduction (measured at 72 weeks)	Mean change in absolute body weight (kg) from Visit 2	MMRM model in Section 4.3.2.2 with body weight at Visit 10 (Week 0) as an additional covariate	LSM estimates through 72 weeks will be plotted by study treatment.
	Mean percent change in body weight from Visit 2	MMRM model in Section 4.3.2.2 with body weight at Visit 10 (Week 0) as an additional covariate	
	Mean change in BMI (kg/m^2) from Visit 2	MMRM model in Section 4.3.2.2 with body weight at Visit 10 (Week 0) as an additional 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
Compare tirzepatide MTD QW with placebo from Visit 2 for waist circumference reduction (measured at 72 weeks)	Mean change in waist circumference (cm)	MMRM model in Section 4.3.2.2	LSM estimates through 72 weeks will be plotted by study treatment.

Abbreviations: ANCOVA = analysis of covariance; BMI = body mass index; HbA1c = hemoglobin A1c; HDL-C = high-density lipoprotein cholesterol; 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; VLDL-C = very low-density lipoprotein cholesterol.

4.5. Exploratory Endpoints Analysis

Table GPHM.4.4. Exploratory Efficacy Analysis

Objective	Relative to the Efficacy Measure	Analysis Conducted
Compare tirzepatide MTD QW with placebo for body weight reduction at 72 weeks	Percentage of participants achieving ≥25% body weight reduction from randomization at 72 weeks	Logistic model in Section 4.3.2.2
Compare tirzepatide MTD QW with placebo for body weight reduction at 72 weeks	Percentage of participants achieving ≥10%, 15%, 20%, and 25% body weight reduction from Week -12 at 72 weeks	Logistic model in Section 4.3.2.2
Assess changes in BMI from Week -12 at Week 0	Percentage of participants whose BMI shifts between clinically relevant categories, that is, from Week -12 (<25, 25 to <30, 30 to <35, 35 to <40, ≥40) to Week 0 (<25, 25 to <30, 30 to <35, 35 to <40, ≥40)	Shift analysis will be conducted based on data from enrolled participants.
Compare tirzepatide MTD QW with placebo for BMI shifts from Week 0 at 72 weeks	Percentage of participants whose BMI shifts between clinically relevant categories, that is, from Week 0 (<25, 25 to <30, 30 to <35, 35 to <40, ≥40) to 72 weeks (<25, 25 to <30, 30 to <35, 35 to <40, ≥40)	Shift analysis will be conducted based on data from the EAS.
Visualize tirzepatide MTD QW and placebo percentage of body weight reduction, from Week -12 up to the safety follow-up after 72 weeks	Percentage of body weight reduction from Week -12 to 72 weeks plus 4 weeks safety follow-up	Time-course plot will be generated. Only the participants who complete the 72 weeks of treatment and have the safety follow-up visit (Visit 801) will be included.
Assess changes from Week -12 at Week 0 and compare tirzepatide MTD QW with placebo in parameters from Week -12 at 72 weeks	Mean change in • systolic blood pressure • diastolic blood pressure • total cholesterol • HDL-C • non-HDL-C • LDL-C • VLDL-C • triglycerides • free fatty acids • fasting glucose • HbA1c • fasting insulin	For analysis from Week -12 at Week 0, change from baseline to postbaseline values at each scheduled visit will be summarized. Within-treatment difference will be assessed by a Wilcoxon signed-rank test. For analysis from Week -12 at 72 weeks, the MMRM model in Section 4.3.2.2 for parameters with repeated measurements will be used. An ANCOVA model for parameters with only 1 postbaseline measurement will be used.
Assess changes in lipid parameters from Week -12 at Week 0	Percentage of participants whose Lipid parameters shift between clinically relevant categories: • LDL-C (<70, 70 to <160, 160 to <190, ≥190 mg/dL) • HDL-C (male: <40, ≥40mg/dL, female: <50, ≥50mg/dL)	Shift analysis will be conducted based on data from enrolled participants.

Approved on 03 Feb 2023 GMT

	• Triglycerides (<150, 150 to <170, 170 to <500, ≥500mg/dL)	
Compare tirzepatide MTD QW with placebo for changes in lipid parameters from Week 0 at 72 weeks	Percentage of participants whose Lipid parameters shift between clinically relevant categories: • LDL-C (<70 mg/dL, 70 to <160 mg/dL, 160 to <190 mg/dL, ≥190 mg/dL) • HDL-C (male: <40 mg/dL, ≥40mg/dL, female: <50 mg/dL, ≥50mg/dL) • Triglycerides (<150 mg/dL, 150 to <170 mg/dL, 170 to <500 mg/dL, ≥500mg/dL)	Shift analysis will be conducted based on data from the EAS.
Assess change in fasting glucose category from Week -12 at Week 0	Percentage of participant change in fasting glucose category, that is, from Week -12 (<100 mg/dL, 100-125 mg/dL, ≥126 mg/dL) to Week 0 (<100 mg/dL, 100-125 mg/dL, ≥126 mg/dL)	Shift analysis will be conducted based on data from enrolled participants.
Compare tirzepatide MTD QW with placebo for change in fasting glucose category from Week 0 at 72 weeks	Percentage of participant change in fasting glucose category, that is, from Week 0 (<100 mg/dL, 100-125 mg/dL, ≥126 mg/dL) to 72 weeks (<100 mg/dL, 100-125 mg/dL, ≥126 mg/dL)	Shift analysis will be conducted based on data from the EAS.
Assess change in HbA1c category from Week -12 at Week 0	Percentage of participant change HbA1c category, that is, from Week -12 (<5.7%, 5.7% to 6.4%, ≥6.5%) to Week 0 (<5.7%, 5.7% to 6.4%, ≥6.5%)	Shift analysis will be conducted based on data from enrolled participants.
Compare tirzepatide MTD QW with placebo for change in HbA1c category from Week 0 at 72 weeks	Percentage of participant change HbA1c category, that is, from Week 0 (<5.7%, 5.7% to 6.4%, ≥6.5%) to 72 weeks (<5.7%, 5.7% to 6.4%, ≥6.5%)	Shift analysis will be conducted based on data from the EAS
Assess change for patient-reported outcome measurements change from Week -12 to Week 0	Mean change in patient-reported outcome measurements (Section 4.7.1)	Data from Week -12 to Week 0 will be included. Change from baseline to postbaseline values at each scheduled visit will be summarized. Within-treatment difference will be assessed by a Wilcoxon signed-rank test.
Compare tirzepatide MTD QW with placebo for patient-reported outcome measurements change from Week 0 to 72 weeks	Mean change in patient-reported outcome measurements (Section 4.7.1)	Data from Week 0 to 72 weeks will be included. An ANCOVA model with terms of treatment (tirzepatide MTD, placebo), country/pooled country, sex, and corresponding baseline score as a covariate will be used. Missing data will be imputed using LOCF.
Compare tirzepatide MTD QW with placebo for patient-reported outcome measurements change	Mean change in patient-reported outcome measurements (Section 4.7.1)	Data from Week -12 to 72 weeks will be included. An ANCOVA model with terms of treatment (tirzepatide MTD, placebo),

from Week -12 to 72 weeks		country/pooled country, sex, and corresponding baseline score as a covariate will be used. Missing data will be imputed using LOCF.

Abbreviations: ANCOVA = analysis of covariance; BMI = body mass index; EAS = Efficacy Analysis Set; HbA1c = hemoglobin A1c; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; LOCF = last observation carried forward; MMRM = mixed model for repeated measures; MTD = maximum tolerated dose; QW = once-weekly; VLDL = very low-density lipoprotein cholesterol.

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

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

4.6. Safety Analyses

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

Safety outcomes during the intensity lifestyle modification lead-in period will also be summarized for all enrolled participants. Detailed analyses can be found in Appendix 4 (Section 6.4).

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

The mean change from baseline differences among treatments at all scheduled visits will be assessed via an MMRM using REML. Unless specified otherwise, if the safety parameter is assessed at 72 weeks, then the model will include country/pooled country, sex, percentage of body weight reduction at the end of lead-in (<10% and ≥10%), 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 do not warrant the MMRM model, then an ANCOVA model will be conducted.

For selected 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 model, assuming the number of events follows a negative binomial distribution, 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 between participants.

Unless otherwise specified, all the analyses listed in this section will be conducted using up to 72 weeks treatment period, plus the safety follow-up (Visit 801), for all randomized participants.

4.6.1. Extent of Exposure

The summary of duration of follow-up (defined as time in days from the date of randomization to the date of the last study visit) and/or duration on study treatment (defined as time in days from the date of first dose of study treatment to the date of last dose of study treatment plus 7 days) will be provided by treatment group using data from the SS over 72 weeks, plus safety follow-up (Visit 801), for all randomized participants.

For the summary of duration on study and study treatment, the frequency and percentage of participants falling into the following ranges 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, and
- ≥72 weeks.

In addition, the frequency and percentages of participants falling into the following study and study treatment exposure ranges 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, and
- ≥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 treatment-emergent adverse event (TEAE) is defined as an event that first occurred, or worsened in severity, after baseline (defined in Table GPHM.4.1). The Medical Dictionary for Regulatory Activities (MedDRA) Lowest-Level Term (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. Events with a missing severity during the baseline period 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 information collected in the case report form (CRF) (for example, treatment-emergent flag, start time of study treatment and event) will be used to determine whether the event was pre- or posttreatment. If the relevant information is not available, then the events will be counted as posttreatment.

In an overview table, the counts and percentages of participants who experienced a TEAE, serious adverse event (SAE), death, or discontinued from study and/or study treatment due to an adverse event (AE) will be summarized by treatment.

The counts and percentages of participants with TEAEs will be summarized by treatment using MedDRA Preferred Terms (PTs) nested within System Organ Class (SOC). Events will be ordered by decreasing frequency within SOC. System Organ Classes will be presented in alphabetical order. Statistical comparisons will be applied at both the SOC and PT levels. If necessary, the percentages of participants with TEAEs will also be summarized by treatment using MedDRA PT (without regard to SOC) by decreasing frequency.

The counts and percentages of participants with TEAEs will be summarized by maximum severity and treatment using MedDRA PTs. 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. 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 Assessments

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, 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 the Clinical Evaluation Committee (CEC), and so on.

4.6.3.2. Other Serious Adverse Events

The counts and percentages of participants who experienced at least 1 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 SOCs will be in alphabetical order.

A listing of all SAEs will be provided. The listing will include treatment, participant identification including the site number, 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 measurements of an individual vital sign (for example, sitting SBP) are collected at the same visit, the mean of these measurements will be used for the analysis.

Vital signs (SBP, DBP, and pulse) will be summarized by treatment group at each planned 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 the analysis model described in Section 4.6. Only planned measurements will be included in the mean change analyses.

The counts and percentages of participants with treatment-emergent abnormal (high or low) vital signs (sitting SBP, DBP, and pulse) at any time during the entire study (including the off-drug follow-up time period) will be summarized by treatment for participants who have both baseline and at least 1 postbaseline result. A treatment-emergent 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 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. Both planned and unplanned measurements will be included in the analysis. The criteria for identifying participants with treatment-emergent vital signs abnormalities are listed in Table GPHM.4.5.

Table GPHM.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

Abbreviation: BP = blood pressure.

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

- counts and percentages of participants who had resting pulse 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 pulse increases from maximum baseline ≥20 bpm at any postbaseline visit
- counts and percentages of participants who had at least 1 resting pulse at any postbaseline visit exceeding 100 bpm, and
- counts and percentages of participants who had at least 1 resting pulse 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 electrocardiogram (ECG) parameters (heart rate, PR, QRS, QT, and QT corrected using Fridericia's correction factor [QTcF] [$QTcF = QT / RR^{0.333}$]). When the QRS is prolonged (for example, a complete bundle branch block), QT and QTc 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 will be summarized for participants who have both a baseline and at least 1 postbaseline result. Treatment differences in mean change from baseline for heart rate and PR interval will be assessed using the analysis model described in Section 4.6. Only planned measurements will be included in the mean change analyses.

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

- treatment-emergent ECG abnormalities as listed in Table GPHM.4.6.
- QT greater than 500 msec

Approved on 03 Feb 2023 GMT

- QTcF greater than 480 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 (see Table GPHM.4.1 for details) 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 GPHM.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: ECG = electrocardiogram; QTcF = Fridericia's corrected QT interval.

4.6.3.6. Clinical Laboratory Evaluation

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

Observed and change from baseline values for each visit may be displayed in plots for participants who have both a baseline and at least 1 postbaseline planned measurement. Baseline is defined in Table GPHM.4.1. Unplanned measurements will be excluded from plots.

A shift table will be provided including unplanned measurements. The shift table will include the number and percentage of participants within each baseline category (minimum value is low, normal, high, or missing) versus each postbaseline category (minimum value is low, normal, high, or missing) by treatment. The Fisher's exact test will be used to compare percentages of participants who shift from normal/high to low between treatments.

For qualitative laboratory analytes, the number and percentage of participants with normal and abnormal values will 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 the 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

The following are "notable" events, from start of study drug through the end of study participation:

- deaths
- SAEs
- permanent discontinuations of study treatment due to AEs, and
- pregnancies.

Patient narratives (participant-level data and descriptive paragraphs) will be provided for participants in the enrolled population with at least 1 notable event.

4.6.5. Special Safety Topics

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

4.6.5.1. 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 values exceeding the following thresholds will be provided by baseline pancreatic enzyme values (≤ upper limit of normal [ULN], > ULN) and postbaseline values (≤1 × ULN, >1 to ≤3 × ULN, >3 to ≤5 × ULN, >5 to ≤10 × ULN, and >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

The counts and percentages of participants with pancreatitis will be presented for both investigator-reported events and subsequently confirmed adjudicated events. Detailed searching criteria can be found in Appendix 7 (Section 6.7).

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

4.6.5.2. Gastrointestinal Adverse Events

4.6.5.2.1. Nausea, Vomiting, and Diarrhea

Summaries and analyses for the incidence and severity of nausea, vomiting (including "vomiting" and "vomiting projectile"), diarrhoea (including "diarrhea" and "diarrhoea"), and the

3 events combined will be provided by each treatment group. A 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 gastrointestinal (GI) AEs (GI SOC) will be captured with the AE-CRF form, and serious cases will be captured with the SAE form. The PTs in the GI SOC from the most recent MedDRA version at the time of DBL will be used to identify GI AEs, and only the PTs with severe/serious cases will be considered AESI.

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

4.6.5.3. Hepatobiliary Disorders

4.6.5.3.1. Hepatic Events

The counts and percentages of participants with treatment-emergent, potentially drug-related hepatic events will be summarized by treatment using MedDRA PTs. Detailed searching criteria can be found in Appendix 7 (Section 6.7). Severe/serious hepatic events will be considered AESI and summarized separately.

4.6.5.3.2. Acute Gallbladder Disease

Events related to acute gallbladder disease will also be summarized by treatment groups by PTs with decreasing frequency. Detailed searching criteria can be found in Appendix 7 (Section 6.7).

Severe/serious acute gallbladder disease will be considered AESI 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 the follow-up period, will be summarized between treatment groups:

- The counts and percentages of participants with an alanine aminotransferase (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 aspartate aminotransferase 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 values, as described above for ALT.

- The counts and percentages of participants with a total bilirubin (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 values:
    - 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 alkaline phosphatase 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 values, 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. Planned and unplanned 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 participant feels that he/she is experiencing symptoms and/or signs associated with hypoglycemia and has a blood glucose (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 (PG) measurements may not be available during such an event, but neurological recovery attributable to the restoration of PG to normal is considered sufficient evidence that the event was induced by a low PG concentration.

Other Hypoglycemia Categories

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

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

Both the incidence (percentage 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.

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.

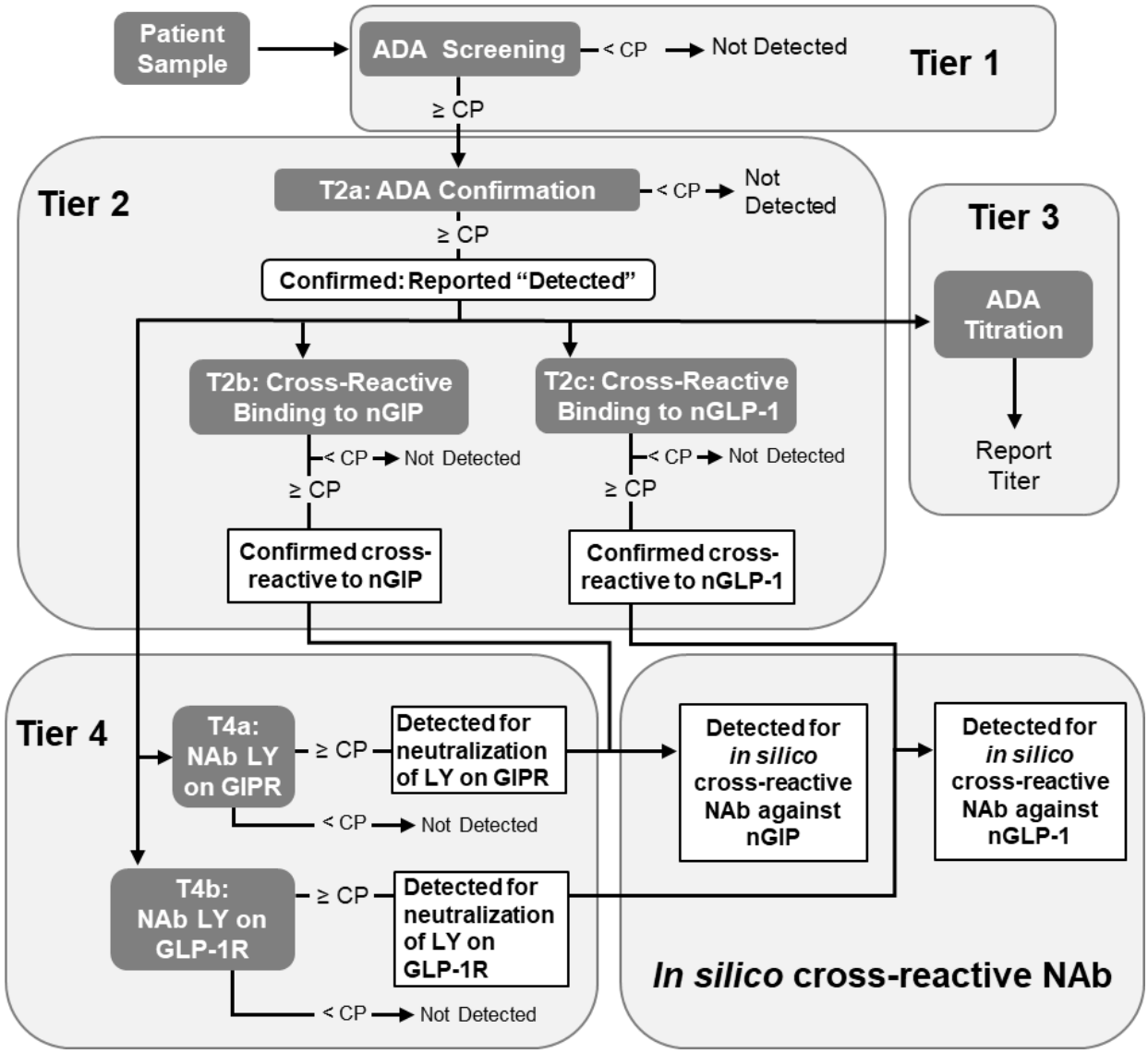
Severe/serious hypoglycemia will be considered AESI. The summaries of severe/serious hypoglycemia will be provided by treatment group. A listing of all events of severe/serious hypoglycemia may be provided, if deemed necessary. This listing will provide treatment allocation and clinical characteristics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

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 anti-drug antibodies (ADA) assay result, potentially multiple cross-reactive ADA assay results, and multiple neutralizing antibodies (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 GPHM.4.1 details a flow chart that reflects the multitiered testing approach.



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

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

Table GPHM.4.7 outlines results as reported from Tier 2a of the multitiered testing approach. Tier 4 results are reported similarly.

Table GPHM.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 GPHM.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 ADA immunoassays, and conversely high levels of ADA 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 GPHM.4.8).

Table GPHM.4.8. Sample Clinical Anti-Drug Antibodies 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 (that is, 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 the drug concentration in the sample is greater than or equal to the assay's drug tolerance level, which may cause interference in the ADA detection method.

Abbreviations: ADA = antidrug antibodies.

All ADA-present samples will be evaluated for cross-reactivity to native glucose-dependent insulinotropic polypeptide (GIP) (Tier 2b), cross-reactivity to native glucagon-like peptide-1 (GLP-1) (Tier 2c), NAb tirzepatide on the GIPR receptor (GIPR) (Tier 4a), and NAb LY (tirzepatide) on the GLP-1 receptor (GLP-1R) (Tier 4b). If cross-reactive ADA against native GIP is detected, the *in silico* assessment for cross-reactive NAb against native GIP is evaluated. If cross-reactive ADA against GLP-1 is detected, the *in silico* assessment for cross-reactive NAb against native GLP-1 is evaluated (Figure GPHM.4.1).

Similar terminology to Table GPHM.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 tirzepatide on GIPR - If an NAb result is not detected and pharmacokinetic (PK) concentration is greater than or equal to the drug tolerance limit of the NAb tirzepatide on the GIPR assay.
- NAb tirzepatide on GLP-1R - If an NAb result is not detected and PK concentration is greater than or equal to the drug tolerance limit of the NAb tirzepatide on the 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 Table GPHM.4.9.

Table GPHM.4.9. *In Silico* Classification for Cross-Reactive NAb

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

In Silico Classification	Cross-Reactive ADA Result	NAb Result	Circulating Tirzepatide Level (ng/mL)	*In Silico* Cross-Reactive NAb Interpretation
	Tier 2c: Detected	Tier 4b: Detected	≥ drug tolerance limit of the assay	Present

Abbreviations: ADA = anti-drug antibodies; NA = not applicable; NAb = neutralizing antibody; 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 GIPR; Tier 4b = NAb LY (tirzepatide) on GLP-1R.

Note that, in the case of an ADA Inconclusive sample, each of the NAb and Cross-Reactive NAb assay results are 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

The baseline period for immunogenicity assessment for each participant includes all observations prior to the first dose of study treatment. In instances where multiple baseline observations are collected, to determine participant ADA status, the last nonmissing immunogenicity assessment prior to first administration of study drug 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:

- 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 the safety follow-up visit or the date of study withdrawal.

4.6.5.5.3. Definitions of Participant ADA Status

Participants Evaluable for Treatment-Emergent ADA (TE ADA)

A participant is evaluable for TE ADA if the participant has a nonmissing baseline ADA result and at least 1 nonmissing postbaseline ADA result.

TE-ADA-Unevaluable Participants

A TE ADA-unevaluable participant is any participant who does not meet the evaluable criteria.

Baseline ADA Present (preexisting antibody)

Anti-drug antibodies are detected in a sample collected up to the first dose date and time.

Baseline ADA Not Present

Anti-drug antibodies are not detected, and the corresponding PK concentrations are missing or are below the drug tolerance limit in a sample collected up to the first dose date and time.

Treatment-Emergent ADA-Positive (TE ADA+) Participants

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 × the minimum required dilution (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 a baseline status of ADA Present, with a titer of 1:baseline, and at least 1 postbaseline status of ADA Present, with a titer of 1:postbaseline, with postbaseline divided by baseline ≥4.

As shown in Figure GPHM.4.1, a titer is expected when an 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 1 dilution above the MRD (1:20).

Treatment-Emergent ADA-Inconclusive Participants

A participant who is evaluable for TE ADA is TE ADA Inconclusive if ≥20% of the participant's postbaseline samples are ADA Inconclusive and all remaining postbaseline samples are ADA Not Present.

Treatment-Emergent ADA-Negative (TE ADA-) Participants

A participant who is evaluable for TE ADA is TE ADA- when the participant is not TE ADA+ and the participant is not TE ADA Inconclusive.

For each NAb assay, the following are defined:

- NAb-positive (NAb+) participant - A participant who is TE ADA+ and has an NAb positive 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 an NAb Inconclusive sample result.
- NAb-negative (NAb-) participant - A participant is NAb- when the participant is neither NAb+ nor 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 participants who are TE ADA evaluable, as defined above. The tabulation will include the number and proportion of participants with

ADA Present at baseline and the number and proportion of TE-ADA+ participants exhibiting each type of cross-reactive antibody and NAb. This analysis will be performed for the following periods:

- 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 a specific TEAE (see Table GPHM.4.10) by participant TE-ADA status (TE ADA+, TE ADA-, or TE ADA Inconclusive).

Table GPHM.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 Activities; SMQ = Standardized MedDRA Query; TEAE = treatment-emergent adverse event.

A listing will be provided for all participants who had ADA Present at any time (including baseline) or had any specific TEAE (see Table GPHM.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 AEs of either severe/serious hypersensitivity or severe/serious injection site reaction (ISR) will be classified as AESI.

Additional immunogenicity analyses, as determined later, may be presented. The relationship between antibody titers, the PK parameters, and pharmacodynamic (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 is 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 the Anaphylactic reaction Standardized MedDRA Query (SMQ) (20000021) (note that the Anaphylactic reaction SMQ algorithm will only be summarized for Time Period A)
- narrow terms in the Angioedema SMQ (20000024)
- narrow terms in the Severe cutaneous adverse reactions SMQ (20000020)
- narrow terms in the Hypersensitivity SMQ (20000214), and
- narrow terms in the 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 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 the 2 time periods:

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

Within queries, individual PTs that satisfied the queries will be summarized. For Time Period A analysis, the Anaphylactic reaction SMQ algorithm will be summarized. Also, a single event may satisfy multiple SMQs, in which case, the event contributes to every applicable SMQ.

4.6.5.6.1. Serious Hypersensitivity Reactions

The severe/serious cases of hypersensitivity will be considered AESI. 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 ISR: erythema, induration, pain, pruritis, and edema.

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

The counts and percentages of participants with treatment-emergent ISRs will be summarized by treatment using MedDRA PTs. Detailed searching criteria can be found in Appendix 7 (Section 6.7).

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 ISRs, based on the TEAE search criteria specified in Appendix 6.7 (Section 6.7), will be considered AESI. The counts and percentages of participants with severe/serious 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 major adverse cardiovascular events (MACE) will be considered AESI:

- 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.

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

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

4.6.5.9. Major Depressive Disorder/Suicidal Ideation

The severe/serious major depressive disorder/suicidal ideation or behavior will be captured as AESI. Adverse events will be searched using MedDRA PTs. Detailed searching criteria can be found in Appendix 7 (Section 6.7).

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 Columbia-Suicide Severity Rating Scale (C-SSRS) and the Patient Health Questionnaire (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 being categorized as follows:

- None (not depressed) - 0 to 4.
- Mild - 5 to 9.
- Moderate - 10 to 14.
- Moderately Severe - 15 to 19.
- Severe - 20 to 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 categories during baseline 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 with at least 1 postbaseline measurement.
- Increase from mild or moderate depression to moderately severe or severe depression - includes participants in the mild or moderate depression category during baseline and 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 a specific plan and intent
- active suicidal ideation with some intent to act without a specific plan
- active suicidal ideation with any methods (no plan) without an intent to act
- nonspecific active suicidal thoughts
- wish to be dead, and
- nonsuicidal, 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 C-SSRS categories by the study participant:

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

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

- **Suicidal ideation** - A "yes" answer at any time during study to any 1 of the 5 suicidal ideation questions (Categories 1 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

Malignancy will be considered as an AESI. The counts and percentages of participants with treatment-emergent malignancy will be summarized by treatment and PT ordered by decreasing frequency. Detailed searching criteria can be found in Appendix 7 (Section 6.7).

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.

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

Mixed model for 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, including those associated with chronic renal failure exacerbation, will also be captured.

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

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

- **Acute renal failure:**
  - narrow terms in the Acute renal failure SMQ (20000003)
- **Chronic kidney disease:**
  - narrow terms in the 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 following SMQ and summarized. Severe/serious dehydration events will be considered AESI:

- narrow terms in the Dehydration SMQ (20000232).

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

4.6.5.12. Thyroid Safety Monitoring

4.6.5.12.1. Calcitonin

The number and proportion 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) and postbaseline value (≤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

Thyroid malignancies and C-cell hyperplasia will be considered AESI. Detailed searching criteria can be found in Appendix 6.7 (Section 6.7).

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

The treatment-emergent arrhythmias and cardiac conduction disorders events will be identified using MedDRA PTs. Detailed searching criteria can be found in Appendix 7 (Section 6.7).

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. Severe/serious treatment-emergent arrhythmias and cardiac conduction disorders will be considered AESI and summarized separately. Treatment of 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 Drug abuse and dependence SMQ (20000101) will be used. Counts and percentages of participants will be summarized by treatment group with decreasing frequency.

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 may 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 for Patient Global Impression of Status for Physical Activity (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 Software 4.5 will be used to derive the following domain and component scores:

- Mental Component Summary (MCS)
- Physical Component Summary (PCS)
- Physical Functioning (PF) domain
- Role-Physical domain (RP)
- Bodily Pain (BP) domain
- 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 score 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 GPHM.4.3 and Table GPHM.4.4.

The empirical cumulative distribution function (eCDF) curves of the change from baseline to 72 weeks in Short Form-36 Version 2 Health Survey acute form (SF-36v2) PF domain scores will be provided by treatment group.

If data allow, analysis for SF-36v2 PF domain scores described in Table GPHM.4.3 will be conducted to evaluate the treatment effect in participants who have limitations in PF at baseline, which are defined as PGIS responses 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 the Impact of Weight on Quality of Life-Lite-Clinical Trials Version (IWQOL-Lite-CT):

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

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

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

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 GPHM.4.3 and Table GPHM.4.4.

If data allow, IWQOL-Lite-CT PF composite score analysis described in Table GPHM.4.3 will be conducted to evaluate the treatment effect in participants who have limitations in PF 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 be conducted:

- descriptive summaries by treatment group, and
- analyses described in Table GPHM.4.4.

4.7.2. Subgroup Analyses

Efficacy subgroup analyses will be guided by the treatment-regimen estimand in the FAS and the efficacy estimand in the 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 (<65, ≥65 years)
- sex (female and male)
- race
- ethnicity
- BMI at randomization (<30, ≥30 and <35, ≥35 and <40, ≥40 kg/m^2)
- region of enrollment (US, outside of US), and
- percentage of body weight reduction at the end of lead-in (<10% and ≥10%).

The outcome measures for the subgroup analyses will include

- percent change in body weight from randomization at 72 weeks, and
- proportion of participants achieving at least 5% body weight reduction from randomization to 72 weeks.

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

Treatment Regimen Estimand

- Conduct the ANCOVA model on the subgroup only with terms of treatment group, country/pooled country, and sex as fixed effects and baseline body weight as a covariate. Missing body weight measurements at 72 weeks will be imputed using the imputation method described in Section 4.3.2.3.
- Full ANCOVA model - Treatment group, subgroup, treatment-by-subgroup-interaction, country/pooled country, and sex as fixed effects and baseline body weight as a covariate. Missing body weight measurements at 72 weeks will be imputed using the imputation method described in Section 4.3.2.3.

Efficacy Estimand

- Conduct the MMRM model on the subgroup only with terms of treatment group, visit, treatment-by-visit-interaction, country/pooled country, and sex as fixed effects, and baseline body weight as a covariate. The variance-covariance structure for within-patient errors will be same as Section 4.3.2.2.
- 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, and sex as fixed effects and baseline body weight as a covariate. The variance-covariance structure for within-patient errors will be same as Section 4.3.2.2.

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

Treatment Regimen Estimand

- Conduct the logistic regression model on the subgroup only with terms of treatment group, country/pooled country, and sex as fixed effects and baseline body weight as a covariate. Missing body weight measurements 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).
- Full logistic regression model - Treatment group, subgroup, treatment-by-subgroup-interaction, country/pooled country, and sex 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).

Efficacy Estimand

- Conduct the logistic regression model on the subgroup with terms of treatment group, country/pooled country, and sex as fixed effects and baseline body weight as covariate. Missing body weight measurement at 72 weeks will be imputed by the predicted value from the 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).

- Conduct the logistic regression model with terms of treatment group, subgroup, treatment-by-subgroup-interaction, country/pooled country, and sex 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 the full MMRM model aforementioned, then the continuous measurements will be dichotomized into status of achieving at least 5% body weight reduction (Yes or No).

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, protocol-planned objectives are changed as below.

Key Secondary Objectives

- added the key secondary objective, "QW tirzepatide MTD is superior to placebo for proportion of participants who achieve ≥20% body weight reduction from randomization to 72 weeks"

Additional Secondary Objectives

- added the additional secondary objective, "Compare QW tirzepatide MTD with placebo for mean change in Non-HDL-cholesterol (mg/dL) from randomization to 72 weeks"

Subgroup Analysis for Primary Analysis

- added the subgroup variable, "region of enrollment (US, outside of US)" for both outcome measures

5. Sample Size Determination

Approximately 1100 participants will be screened, and 800 participants will be enrolled into the 12-week lead-in period in order to achieve approximately 600 participants randomly assigned to study drug intervention (300 participants per treatment group).

The sample size determination assumes that evaluation of superiority of tirzepatide MTD (10 mg or 15 mg) to placebo will be conducted at a 2-sided significance level of 0.05 using a 2-sample t-test. Additionally, a difference of at least 12% in mean body weight percentage reduction from randomization at 72 weeks for tirzepatide MTD 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 600 participants in a 1:1 ratio to MTD (300 participants) and placebo (300 participants) provides more than 90% power to demonstrate superiority of tirzepatide MTD to placebo.

The chosen sample size and randomization ratio also provides >90% power to establish superiority of tirzepatide MTD to placebo in terms of proportion of participants achieving at least 5% body weight reduction from randomization at 72 weeks, based on a Chi-square test at a 2-sided significance level of 0.05, assuming 20% of placebo-treated participants and 46% of tirzepatide-treated participants achieve the goal and a dropout rate of 25%.

Approved on 03 Feb 2023 GMT

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), country, and weight-related comorbidities.

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 MedDRA PTs nested within SOC. The SOCs will be in alphabetical order. Conditions (that is, PTs) will be ordered by decreasing frequency within SOC. This will be summarized for all randomized participants.

6.1.3. Concomitant Therapy

Concomitant medications will be summarized by PTs by treatment group by decreasing frequency for the SS group. The following postbaseline period will be considered:

- up to 72 weeks plus the safety follow-up (Visit 801) for all randomized participants.

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

Concomitant medications of interest include the following:

- baseline antihypertensive therapy, by type/class
- baseline lipid lowering therapy, by type/class
- changes to baseline medication postrandomization
    - antihypertensive therapy, and
    - lipid lowering therapy.
- utilization after randomization of
    - medicines that cause weight gain,
    - antihyperglycemic medication for the treatment of diabetes for participants who develop T2DM during the study (antihyperglycemic medication for the treatment of prediabetes is not allowed per protocol)
    - antidiarrheal medication, and
    - antiemetic medication.

The analysis of concomitant medication for the lead-in period is provided in Appendix 4 (Section 6.4).

6.2. Appendix 2: Treatment Compliance

A summary of participants who prematurely discontinued 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.

The analyses related to compliance during the 72-week treatment period will be conducted for all randomized participants.

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

Treatment compliance will be defined as taking at least 75% of the scheduled study treatment doses. Compliance over the double-blind treatment period will be calculated using the number of doses administered (regardless of the actual dose in mg administered), divided by the total number of doses expected to be administered × 100 over the double-blind treatment period, respectively. Treatment compliance will be summarized descriptively over the double-blind treatment period by treatment using the mITT population.

6.3. Appendix 3: Important Protocol Deviations

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

6.4. Appendix 4: Analyses for the Intensive Lifestyle Modification Lead-in Period

6.4.1. General Consideration

Unless specified otherwise, all analyses for the lead-in period will be based on all enrolled participants.

Table GPHM.6.1. Baseline and Postbaseline Definitions in Lead-in Period

Population	Analysis Type	Baseline	Postbaseline
All enrolled participants	1.1) Treatment-emergent adverse events	The baseline period is defined as the start of screening and ends prior to the intensive lifestyle modification program (typically at Week -12).	Starts after the first visit of the intensive lifestyle modification program and ends prior to the first dose of study treatment.

Population	Analysis Type	Baseline	Postbaseline
All enrolled participants	1.2) Treatment-emergent abnormal labs, vital signs, and ECGs.	Baseline will include all scheduled and unscheduled measurements during the baseline period as defined above (1.1).	Postbaseline will be defined as above (1.1). All scheduled and unscheduled measurements will be included.
All enrolled participants	1.3) Change from last baseline to Week *xx* and to last postbaseline for labs, vital signs, and ECGs.	The last scheduled and unscheduled nonmissing assessment recorded during the baseline period defined above (1.1)	Postbaseline will be defined as above (1.1). Only scheduled visits will be included. The early discontinuation visits are considered scheduled visits.
All enrolled participants	Other outcomes (efficacy, PROs, questionnaires, etc.)	Baseline is defined as the last nonmissing data collected prior to or at study entry (Week -12).	Postbaseline is defined as data collected after the study entry (Week -12) and prior to or at randomization (Week 0)

Abbreviations: ECG = electrocardiogram; PRO = patient-reported outcome.

6.4.2. Patient Characteristics

A summary table will be generated for participant demographics and baseline disease characteristics for the enrolled population. Variables to be included (but not limited to) are age, sex, race, ethnicity, weight, BMI, waist circumference, age groups, and BMI groups.

6.4.3. Disposition

Summaries of study disposition will be provided for the intensive lifestyle modification lead-in period. A listing may also be provided if necessary.

The lead-in period completion status is defined at the end of 12-week intensive lifestyle modification lead-in period (when all enrolled participants complete or discontinue the lead-in period). Participants with a nonmissing body weight measurement at Week 0 (Visit 10) will be considered as completers; otherwise, they will be considered as noncompleters.

6.4.4. Medical History, Preexisting Conditions, and Concomitant Medications

Medical history, preexisting conditions, and concomitant medications will be summarized by PTs by decreasing frequency for all enrolled participants.

6.4.5. Adverse Events

For the intensive lifestyle modification lead-in period, a TEAE is defined as an event, based upon an MedDRA LLT, that first occurred or worsened in severity after baseline of the lead-in period.

Treatment-emergent AEs, common TEAEs, deaths, SAEs, and significant AEs will be summarized for all enrolled participants in a similar way to Section 4.6.2 without treatment comparison.

6.4.6. Additional Safety Assessments

Vital signs, ECG parameters and clinical laboratory results will be summarized for all enrolled participants using categorical analysis in a similar way to Section 4.6.3 without treatment comparison.

6.4.7. Important Protocol Deviations

Important protocol deviations are identified in the TIMP. A summary of important protocol deviations may be provided at the end of the 12-week intervention (for all enrolled participants), if deemed necessary.

6.5. Appendix 5: Clinical Trial Registry Analyses

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

Analyses provided for the CTR requirements include the following:

- Summary tables of AEs for the lead-in period and the double-blind period, provided as datasets that will be converted to XML files. For the lead-in period, SAEs and 'other' nonserious AEs are summarized by MedDRA PT. For the double-blind period, they are 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 SAE 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.

A demographic table including the following age ranges is required by EudraCT:

- 18 to 65 years
- 65 to 85 years, and
- 85 years and over.

6.6. Appendix 6: Exceptional Circumstances Impact

This section lists additional statistical analyses that may be performed at final database lock to assess the impact of exceptional circumstances if the data warrants.

6.6.1. General Considerations

Percentage and count of randomized participants who followed the exceptional circumstances 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, and so on.

Percentage and count of randomized participants who completely missed at least 1 study visit due to exceptional circumstances may also be summarized.

Similar summaries during the intensive lifestyle modification lead-in period may also be provided for all enrolled participants.

6.6.2. Exposure

A listing of randomized participants who had the study drug temporarily interrupted due to exceptional circumstances may be provided.

6.6.3. Protocol Deviations

Percentage and count of randomized participants having important protocol deviations related to exceptional circumstances will be summarized by treatment.

Percentage and count of randomized participants with protocol deviation and mitigation related to exceptional circumstances may also be summarized by treatment.

A listing of all randomized participants who had protocol deviation and mitigation due to exceptional circumstances may be provided.

Similar summaries during the intensive lifestyle modification lead-in period may also be provided for all enrolled participants.

6.6.4. Patient Disposition

A summary table for all randomized participants that discontinue the study or study treatment due to exceptional circumstances will be provided by treatment. A similar summary during the intensive lifestyle modification lead-in period may also be provided for all enrolled participants.

A listing of enrolled participants who discontinued the study or study treatment due to exceptional circumstances will be provided with information to indicate if a participant is randomized (along with randomized treatment) and if events occur postrandomization.

6.6.5. Adverse Events

A listing of all enrolled participants who had COVID-19 infection, including deaths due to COVID-19, will be provided with information to indicate if a participant is randomized (along with randomized treatment), and if events occur post-randomization.

6.6.6. 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 subgroup impacted

by exceptional circumstances (that is, participants without impact versus with impact) for the SS group.

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

Similar summaries during the intensive lifestyle modification lead-in period may also be provided for all enrolled participants.

6.6.7. Local Laboratory Measurements

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

6.6.8. 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.

6.7. Appendix 7: Searching Criteria for Special Safety Topics

Pancreatitis Events

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

Malignancies

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

Arrhythmias and Cardiac Conduction Disorders

The cardiovascular events will include clinically relevant rhythm and conduction disorders. The treatment-emergent arrhythmias and cardiac conduction disorders events will be identified using the MedDRA PT contained in any of the following SMQs:

1) Arrhythmias
    - For symptoms - Arrhythmia related investigations, signs and symptoms SMQ (20000051), narrow and broad terms.
    - For supraventricular arrhythmias - Cardiac arrhythmia SMQ, under the tachyarrhythmia sub-SMQ.
        - Supraventricular tachyarrhythmia SMQ (20000057), broad and narrow terms

- Tachyarrhythmia terms, nonspecific SMQ (20000164), narrow terms only, and
- Ventricular tachyarrhythmia SMQ (20000058), narrow terms only.

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

Injection Site Reactions

Treatment-emergent ISRs will be identified using the MedDRA PT in any of the following

- MedDRA HLT of Injection site reaction
- HLT of Administration site reaction, and
- HLT of Infusion site reactions.

Acute Gallbladder Disease

All events of TEAE biliary colic, cholecystitis, or other suspected events related to gallbladder disease will be identified by PTs under following SMQs:

- narrow PTs in the Gallbladder related disorders SMQ (20000124)
- narrow PTs in the Biliary tract disorders SMQ (20000125), and
- narrow PTs in the Gallstone related disorders SMQ (20000127).

Hepatic Events

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

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

Major Depressive Disorder/Suicidal Ideation

Adverse events will be searched using MedDRA PTs from the Depression and suicide/self-injury SMQ as defined in MedDRA (SMQs: 20000037 [Suicide/self-injury] and 20000167 [Depression (excl suicide and self-injury)]).

7. References

[ADA] American Diabetes Association. Glycemic targets: standards of medical care in diabetes-2020. *Diabetes Care.* 2020(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, Meyer DR. 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;21:641-653. 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-076448 v1.0

Approval	PPD 03-Feb-2023 19:44:23 GMT+0000

Signature Page for VV-CLIN-076448 v1.0