Statistical Analysis Plan: I8F-MC-GPID

A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study Comparing the Efficacy and Safety of Tirzepatide versus Placebo in Patients with Heart Failure with Preserved Ejection Fraction and Obesity (SUMMIT)

NCT04847557

Approval Date: 26-Jun-2024

Title Page

Protocol Title: A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study Comparing the Efficacy and Safety of Tirzepatide versus Placebo in Patients with Heart Failure with Preserved Ejection Fraction and Obesity (SUMMIT)

Protocol Number: I8F-MC-GPID

Compound Number: LY3298176

Short Title: A Statistical Analysis Plan for Tirzepatide in Participants with Heart Failure with Preserved Ejection Fraction and Obesity (SUMMIT)

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

Registry	ID
IND	147352

Document ID: VV-CLIN-064990

Table of Contents

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Version history

This is the second version of the statistical analysis plan (SAP) for Study I8F-MC-GPID (GPID), which is based on Protocol (c) GPID, approved on 14 February 2024. SAP GPID version 1 was approved on 22 November 2021. See approval date for the current version of this SAP on Page 1.

Major Revision Summary for I8F-MC-GPID Statistical Analysis Plan Version 2

Section # and Name	Description of Change	Brief Rationale
Section 1.1 Primary Objective	Revised primary endpoints to • change from baseline to Week 52 in the KCCQ-CSS, and • occurrence of the composite endpoint of CV death and/or HF events over time. Hierarchical composite assessed by win ratio moved to other secondary endpoint. Change from baseline to Week 52 in 6MWD moved to Key Secondary endpoints.	To align with protocol (c).
Section 1.2 Key Secondary Objectives	Key secondary endpoints revised to • change from baseline to Week 52 in 6MWD • percent change from baseline to Week 52 in body weight, and • change from baseline to Week 52 in hsCRP. NYHA class, exercise capacity (6MWD at Week 24) moved from Key Secondary endpoints to Other Secondary endpoints.	To align with protocol (c)

Section # and Name	Description of Change	Brief Rationale
Section 3.2 General consideration	Change end of study participation to end of follow-up. Add detail of time to event derivation.	To provide more details of the updated primary analysis and to align with the new primary endpoints.
Section 3.11.1 Study and Study Treatment Exposure	Included longer follow-up interval when summarizing duration.	To provide more details of exposure.
Section 3.11.2 Compliance to Study Treatment	Clarification of compliance criteria.	To align with protocol (c)
Section 3.13.1 Primary Endpoints/Estimands Analysis	Revised statistical methods related to primary endpoints.	To align with the revised primary endpoint and provide an appropriate analysis method for each endpoint/estimand.
Section 3.13.2 Key Secondary Endpoints/Estimands	Revised statistical methods related to key secondary endpoints.	To align with the revised primary endpoint and provide an appropriate analysis method for each endpoint/estimand.
Section 3.13.3 Type I Error Rate Control Strategy for Primary and Key Secondary Efficacy Analyses	Updated type I error rate control strategy	To provide details of the strategy.
Section 3.13.4 Other Secondary	Other Secondary endpoints added including: • hierarchical composite assessed by win ratio • clinical outcome events of HF • NYHA Class • proportion of participants attaining ≥5%, ≥10%, ≥15% and ≥20% in body weight reduction change at Week 52	To align with the revised primary endpoint and provide an appropriate analysis method for each endpoint/estimand.

Approved on 26 Jun 2024 GMT

Section # and Name	Description of Change	Brief Rationale
	• change from baseline to Week 24 in KCCQ-CSS added to other secondary endpoints • proportion of participants attaining KCCQ-CSS MWPC threshold at Week 52 added to other secondary endpoints • proportion of participants attaining ≥5 meters, ≥10 meters, and ≥15 points KCCQ-CSS change at Week 52 • change from baseline to Week 24 in 6MWD • proportion of participants attaining 6MWD meaningful within-patient change (MWPC) threshold at Week 52, and • proportion of participants attaining ≥10 meters, ≥20 meters, and ≥30 meters 6MWD change at Week 52.	
Section 3.13.5 Exploratory Endpoints	Exploratory endpoint "HF medication use" integrated into primary endpoint CV death and/or HF event. Clinical outcome events of HF moved to other secondary. Exploratory endpoint "Evaluation of prespecified	To align with the revised primary endpoint and provide an appropriate analysis method for each endpoint/estimand.

Section # and Name	Description of Change	Brief Rationale
	biomarkers" hsCRP; moved to key secondary endpoint. Added exploratory endpoints: • change from baseline to week 52 in waist to height ratio, and • eGFR slope.	
Section 3.14.3 Special Safety Topics	Added safety topics.	To align with PSAP GPID version 4.
Section 3.14.3.5.2 Liver Enzymes	Updated categorization.	To align with the new Lilly hepatic analyses plan.
Section 3.15 Subgroup Analysis	Updated categorization of subgroup variables. Removed subgroup of ARNi and age category of 75 and added new subgroups. Added safety in special groups and situations section.	To be consistent with baseline characteristics and to provide additional subgroups analysis. To align with PSAP GPID version 4 and provide submission level safety analyses details.
Section 5.1, Appendix 1	Added Kansas City Cardiomyopathy Questionnaire.	To provide details of the questionnaire.
Section 5.2, Appendix 2	Added further detailed search criteria for analysis of special safety topics.	To align with PSAP GPID version 4 and provide detailed descriptions of MedDRA search criteria.
Section 5.3, Appendix 3	Added additional endpoints of interest and analysis details.	To provide additional details of analysis for the cardiac MRI substudy.
Section 5.4, Appendix 4	Added statistical analysis for China.	To specify analyses to be performed for participants enrolled in mainland China and Taiwan.

Abbreviations: 6MWD = 6-minute walk test distance; ARNi = angiotensin receptor-neprilysin inhibitors; CV = cardiovascular; eGFR = estimated glomerular filtration rate; HF = heart failure; hsCRP = high-sensitivity C-reactive protein ; KCCQ-CSS = Kansas City Cardiomyopathy Questionnaire-Clinical Summary Score; MedDRA = Medical Dictionary for Regulatory Activities; MWPC = meaningful within-patient change; NYHA = New York Heart Association; PSAP = program statistical analysis plan.

1. Study Objectives

1.1. Primary Objective

To demonstrate that a maximally tolerated dose (MTD) of tirzepatide up to 15 mg administered subcutaneously (SC) once weekly (QW) is superior to placebo in participants with heart failure (HF) with preserved ejection fraction (HFpEF) and body mass index (BMI) ≥30 kg/m^2 based on:

- change from baseline in Kansas City Cardiomyopathy Questionnaire (KCCQ) Clinical Summary Score (CSS) at Week 52, and
- occurrence of the composite endpoint of cardiovascular (CV) death and/or heart failure (HF) events over time.

The effectiveness of tirzepatide will be demonstrated if either one or both the primary objectives is met. The details on type I error control are discussed in Section 3.13.3.

The HF event definition within Protocol GPID includes worsening symptoms or signs of HF, which are meaningful to the participant and require intensification of treatment characterized by one or more of the following: hospitalization for heart failure regardless of duration or treatment received; use of intravenous drug, usually an intravenous diuretic, but may include intravenous vasodilators or positive inotropic drugs; or augmentation or increase in oral diuretic therapy.

1.2. Key Secondary Objectives

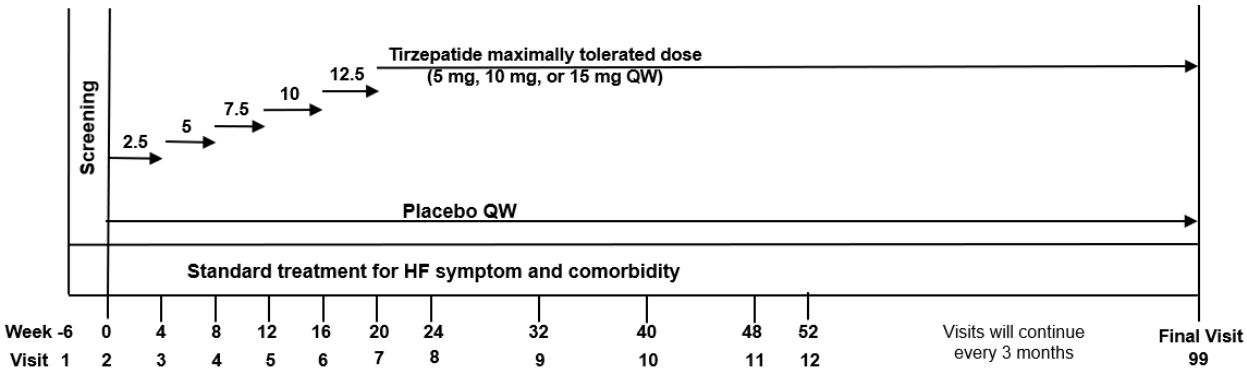
The key secondary objectives are to demonstrate that tirzepatide MTD is superior to placebo with regards to

- change from baseline in 6-minute walk test distance (6MWD) at Week 52
- percent change from baseline in body weight at Week 52, and
- change from baseline in high-sensitivity C-reactive protein (hsCRP) at Week 52.

All the key secondary objectives are under multiplicity control. The details on type I error control are discussed in Section 3.13.3.

2. Study Design

Study GPID is a randomized, outpatient, multicenter, international, placebo-controlled, double-blinded, parallel arm Phase 3 study with 2 study periods. The study is designed to evaluate the efficacy and safety of QW tirzepatide MTD up to 15 mg compared to placebo in participants with HFpEF and BMI ≥30 kg/m2. Figure GPID.2.1 illustrates the Study GPID design.



Note: Screening procedures may take longer or shorter than 6 weeks and variation in screening procedures will not be considered a protocol deviation.

Figure GPID.2.1. Illustration of study design for Clinical Protocol I8F-MC-GPID.

Two intervention groups will be studied:

- tirzepatide MTD up to 15 mg SC QW, and
- placebo.

Study GPID will compare treatment with tirzepatide and treatment with placebo. Assignment to tirzepatide or placebo groups will be randomly allocated in a 1:1 ratio.

The starting dose of tirzepatide is 2.5 mg QW, which will be escalated at 4-week intervals to a maximum of 15 mg QW, or to the highest maintenance dose tolerated by the participant.

The study will consist of 2 periods:

- Study Period 1: screening period, up to approximately 6 weeks, and
- Study Period 2: treatment period, with a 20-week escalation followed by at least a 32-week maintenance period.

Study GPID will continue until approximately 52 weeks after the last participant is randomized. The maximum duration of an individual's participation will depend on the duration of study enrollment.

3. A Priori Statistical Methods

3.1. Populations for Analyses

The populations for analyses are defined in the following table (Table GPID.3.1).

Table GPID.3.1. Description of Analysis Populations

Analysis Population	Description
Entered	All participants who sign the ICF.
Randomized/ITT Population	All participants assigned to treatment, regardless of whether they take any doses of study treatment, or if they took the correct treatment. Participants will be analyzed according to the treatment group to which they were assigned.
Safety Population	All participants in the ITT population who take at least 1 dose of study treatment. Participants will be analyzed according to the treatment group to which they were assigned.

Abbreviations: ICF = informed consent form; ITT = intent-to-treat.

3.2. General Considerations

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

Statistical treatment comparisons will be performed between tirzepatide MTD and placebo. Unless otherwise specified, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and all confidence intervals (CIs) will be given at a 2-sided 95% level. Efficacy will be assessed using the intent-to-treat (ITT) population. Baseline is defined as the last nonmissing observation collected prior to or at randomization for efficacy analyses. Safety will be assessed using safety population and the definition of baseline and postbaseline for safety analyses are specified in Table GPID.3.2.

Table GPID.3.2. Baseline and Postbaseline Definitions for Safety Analyses

Analysis Type	Baseline	Postbaseline
1.1) Treatment-Emergent Adverse Events	The baseline period is defined as the start of screening and ends prior to the first dose of study treatment (typically at Week 0).	Starts after the first dose of study treatment and ends at the end of follow-up. See Section 3.2 for the definition of a participant's end of follow-up.
1.2) Treatment-Emergent Abnormal Labs[a] and Vital Signs	Baseline will include all scheduled and unscheduled measurements during the baseline period (Visit 1 and Visit 2).	Postbaseline will be defined as measurements after Visit 2. All scheduled and unscheduled measurements will be included.
Analysis Type	**Baseline**	**Postbaseline**
1.3) Change from Last Baseline to Week xx and to Last Postbaseline for Labs[a] and Vital Signs	The last scheduled and unscheduled nonmissing assessment recorded during the baseline period defined above (1.2).	Postbaseline will be defined as above (1.2). Only scheduled visits will be included. The ET visits are considered scheduled visits.

Abbreviations: ET = early termination.

[a] Immunogenicity related analysis is specified in Section 3.14.3.8.

Summary descriptive statistics for continuous measures will include sample size, mean, standard deviation (SD), median, minimum, and maximum. Summary statistics for categorical measures (including categorized continuous measures) will include sample size, frequency, and percentages. Summary statistics for discrete count measures will include sample size, mean, SD, median, minimum, and maximum.

Unless otherwise specified, all observed data will be considered for analysis regardless of adherence to randomized treatment. The participant's end of follow-up will be the later of last contact date or study disposition date.

The definition of time-to-event for a specific event of interest is specified in Table GPID.3.3.

Table GPID.3.3. Definition of Time-to-Event for a Specific Event of Interest

If a participant:	then:
Experiences the event	time-to-event for a specific event of interest will be the number of days between the date of randomization and the onset date of the event plus 1 day.
Does not experience the event	the participant will be censored and the number of days between the date of randomization and the date of the participant's end of follow-up plus 1 day will be used for analysis.

Not all analyses described in SAP GPID will necessarily be included in the clinical study report (CSR). Any analysis described in this SAP and not provided in the CSR will 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.

3.3. Adjustments for Covariates

The study is stratified by diagnosed type 2 diabetes mellitus (T2DM) (Y/N), HF decompensation (including hospitalization for HF requiring intravenous [IV] diuretic treatment or urgent HF visit requiring IV diuretic treatment) within 12 months of screening (Y/N), and baseline BMI group (<35, ≥ 35 kg/m^2). Unless otherwise specified, the stratification factors will be adjusted in the efficacy analyses. The value for stratification factors will be obtained from the data collected or derived from the electronic case report form (eCRF). In addition, the baseline value of the endpoint will be used as a covariate when appropriate.

3.4. Handling of Dropouts or Missing Data

For the primary and key secondary efficacy endpoint analyses subject to type I error rate control, missing data will be imputed based on the methods described in Section 3.13.1.3 and 3.13.2.

For all other endpoints, missing values will not be explicitly imputed unless specified otherwise.

3.5. Multicenter Studies

There is no stratification by site or country for randomization. However, the country or region effect may be examined for the primary endpoints through subgroup analysis.

3.6. Multiple Comparisons/Multiplicity

The type I error rate control strategy for primary and key secondary objectives is discussed in Section 3.13.3. There will be no multiplicity adjustments for evaluating other secondary and exploratory objectives and safety assessments.

3.7. Patient Disposition

Summaries and a listing of study disposition and study drug disposition will be provided for all randomized participants. Comparison between treatment arms will be performed using Fisher's exact test.

Summaries of study disposition will be provided for all entered but not randomized participants.

3.8. Historical Illnesses and Preexisting Conditions

The count and percentages of participants with historical illnesses and preexisting conditions will be summarized by treatment group using the Medical Dictionary for Regulatory Activities (MedDRA Version 27.0) preferred terms (PTs) nested within system organ class (SOC). The SOC will be in alphabetical order. Conditions (that is, PTs) will be ordered by decreasing frequency in tirzepatide MTD arm within SOC. This will be summarized for all randomized participants. Historical illnesses and preexisting conditions of special interest will also be summarized separately.

3.9. Patient Characteristics

A listing of participant demographics for all randomized participants will be provided. The demographic and baseline clinical characteristics will also be summarized by study treatment

group 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, ≥65; <75, ≥75; <65, ≥65 and <75, ≥75 and < 85, ≥85)
- BMI group (≥30 and <35, ≥35 and <40, ≥40 kg/ m^2)
- country
- vital signs, and
- characterization of HFpEF and HFpEF-related comorbidities.

3.10. Concomitant Therapy

Concomitant medication will be summarized by treatment groups and displayed by decreasing frequency of WHODrug (Version MAR24B3) PTs in tirzepatide MTD arm.

In addition, medications of interest (as defined below) will be summarized by treatment groups:

- Baseline use of
  - HF medications
  - antihypertensive therapy other than HF medications
  - antiplatelet and anticoagulant medications
  - antihyperglycemic medications, and
  - lipid lowering therapy.
- Changes to baseline medication post-randomization
  - HF medications
  - antihypertensive therapy other than HF medications
  - antihyperglycemic therapy, and
  - lipid lowering therapy.

3.11. Treatment Exposure and Compliance

3.11.1. Study and Study Treatment Exposure

A summary of duration on study follow-up (defined as time in days from date of randomization to the date of the end of follow-up plus 1 day) will be provided by treatment group in the ITT population.

A summary of duration on study treatment (defined as time in days from date of first dose of study treatment to date of last dose of study treatment plus 7 days) will be provided by treatment group in the safety population.

For the summary of duration on study and study treatment, the frequency and percentage of participants falling into the following categorical ranges will also be summarized by planned treatment group: >0 week, ≥4 weeks, ≥8 weeks, ≥12 weeks, ≥16 weeks, ≥20 weeks, ≥24 weeks, ≥32 weeks, ≥40 weeks, ≥48 weeks, ≥52 weeks, ≥65 weeks, ≥78 weeks, ≥91 weeks, ≥104 weeks, ≥117 weeks, ≥130 weeks, ≥143 weeks, ≥156 weeks.

In addition, the frequency and percentages of participants falling into the following exposure ranges for study and study treatment may be summarized by planned treatment group: >0 to <4 weeks, ≥4 to <8 weeks, ≥8 to <12 weeks, ≥12 to <16 weeks, ≥16 to <20 weeks, ≥20 to <24 weeks, ≥24 to <32 weeks, ≥32 to <40 weeks, ≥40 to <48 weeks, ≥48 to <52 weeks, ≥52 weeks to <65 weeks, ≥65 weeks to <78 weeks, ≥78 weeks to <91 weeks, ≥91 weeks to <104 weeks, ≥104 weeks to <117 weeks, ≥117 weeks to <130 weeks, ≥130 weeks to <143 weeks, ≥143 weeks to <156 weeks, ≥156 weeks.

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

3.11.2. Compliance to Study Treatment

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

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

Treatment compliance will be defined as taking at least 75% of the scheduled tirzepatide doses. Similarly, a participant will be considered significantly noncompliant if he or she is judged by the investigator to have intentionally or repeatedly taken more than the prescribed amount of medication (more than 125%).

Compliance over the whole 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 treatment period. Total number of doses expected is defined as the number of weeks between the treatment end date and first dose date minus the number of dose interruptions due to adverse event (AE), investigator decision, or abnormal lab results if there are any. Treatment compliance will be summarized descriptively over the treatment period by treatment using the safety population.

3.12. Important Protocol Deviations

Important protocol deviations are defined in the Trial Issues Management Plan. A listing and a summary of important protocol deviations by treatment will be provided.

3.13. Efficacy Analyses

The primary estimand for primary endpoints and key secondary endpoints is to assess the treatment difference between tirzepatide and placebo relative to the efficacy measures for all randomized participants. The treatment policy strategy will be used to handle intercurrent events (ICEs), meaning all the observed values for the variable of interest are used regardless of whether or not the ICE occurs. The details of the primary estimand for each endpoint will be described in the following sections.

3.13.1. Primary Endpoints/Estimands Analysis

3.13.1.1. Change from Baseline in Kansas City Cardiomyopathy Questionnaire-Clinical Summary Score (KCCQ-CSS)

3.13.1.1.1. Estimand for the Endpoint

The estimand for the primary endpoint is described by the following attributes:

- Endpoint: change from baseline to Week 52 in KCCQ-CSS.
- Treatment condition of interest: tirzepatide MTD versus placebo.
- Handling of ICEs: using the treatment policy strategy to handle all ICEs.
- Population: all randomized participants.
- Population-level summary: the median difference between treatment arms.

The KCCQ is a 23-item, participant self-administered questionnaire that assesses impacts of HF "over the past 2 weeks" on 7 domains (Green et al. 2000; Joseph et al. 2013). Each of the 23 individual items are answered on Likert scales of varying length (5-point, 6-point, or 7-point scales). Domain scores are obtained by averaging the associated individual items and transforming the score to a 0 to 100 range. Higher scores indicate better health status. KCCQ-CSS is a summary score that is derived by taking the mean of the Physical Limitation and Total Symptom scores. Detailed scoring instructions are provided in Appendix 1, Section 5.1.

3.13.1.1.2. Main Analytical Approach

A stratified Wilcoxon test (van Elteren 1960) will be used as the main analysis method, controlling for the stratification factors of HF decompensation within 12 months of screening (Y/N), diagnosed T2DM (Y/N), and baseline BMI (<35, ≥35 kg/m^2). The Hodges-Lehmann estimate for the median difference and 2-sided 99% and 95% CIs will be reported.

Missing KCCQ-CSS measurements at Week 52 will be imputed through multiple imputation as specified in Section 3.13.1.3. The complete datasets generated through multiple imputation will be analyzed and a Van Elteren test will be conducted for treatment comparison. The final statistical inference over multiple imputation will be guided by the method proposed by Rubin (1987).

The empirical cumulative distribution function and histogram of the change from baseline to Week 52 in the KCCQ-CSS will be provided by treatment group.

3.13.1.2. Occurrence of Cardiovascular Death and/or Heart Failure Event over time

3.13.1.2.1. Time to first occurrence of Cardiovascular Death or Heart Failure Event

3.13.1.2.1.1. Estimand for the Endpoint

The estimand for the primary endpoint is described by the following attributes:

- Endpoint: time from randomization to the first occurrence of the clinical endpoint committee (CEC)-confirmed heart failure events or CV death.
- Treatment condition of interest: tirzepatide MTD versus placebo.
- Handling of ICEs: using treatment policy strategy to handle all ICEs.
- Population: all randomized participants.
- Population-level summary: hazard ratio.

3.13.1.2.1.2. Main Analytical Approach

The primary analysis will be a Cox proportional hazards model with treatment as a fixed effect adjusting for diagnosed T2DM (Y/N), baseline probability of HFpEF (<0.8, ≥0.8), and baseline N-terminal pro b-type natriuretic peptide (NTproBNP) (<200, ≥200 ng/L). The probability of HFpEF is derived from the HFpEF-ABA model (Reddy et al 2024). Participants who did not have an adjudicated primary endpoint event on or prior to the end of follow-up will be censored at the date of participant's end of follow-up. The missing data due to censoring will be implicitly handled by the Cox regression model, assuming censoring is independent of the outcome. The hazard ratio, with its CI and p-value, will be provided through the primary analysis model.

The Kaplan-Meier method will be used to estimate the cumulative event curve over time. Counts and proportions of participants who experience a primary endpoint event will be calculated. The total person-years of follow-up for the primary endpoint, the incidence rate per 100 person-years of follow-up for the primary endpoint, and the absolute risk difference (ARD) for the primary endpoint will be provided.

Person-years of follow-up for a specific event of interest are defined for each participant as the time-to-event divided by 365.25.

The incidence rate per 100 person-years of follow-up is defined by dividing the number of participants who developed the event during the study period by the event specific total person-years of follow-up (that is, time-to-event as defined above) multiplied by 100.

The ARD for an endpoint is defined as the difference in incidence rate per 100 person-years between the 2 treatment groups (placebo minus tirzepatide).

3.13.1.3. Methods for Missing Data Imputation

The missing measurement for KCCQ-CSS at 52 weeks for the primary estimand will be imputed through multiple imputation based on the reason for missingness.

- For missing measurements due to death, multiple imputation will be performed using the worst 15% observed data at 52 weeks from the same treatment group.
- For missing data due to all other ICEs or without ICE, retrieved dropout imputation will be applied, which will utilize observed data from participants in the same treatment group

who had outcome measures at Week 52 after early discontinuation of study drug to impute the missing value. In case there are not enough retrieved dropouts to provide a reliable imputation model, reference to the placebo imputation will be used.

3.13.1.4. Sensitivity Analyses for Kansas City Cardiomyopathy Questionnaire-Clinical Summary Score

For the primary endpoint of change from baseline in KCCQ-CSS, a mixed-effects model repeated measures (MMRM) analysis will be conducted to analyze the change from baseline in the KCCQ-CSS. The analysis will be guided by the treatment policy strategy. All values of the collected KCCQ-CSS data at baseline, 24 weeks, and 52 weeks will be used in the MMRM analysis. The primary endpoint assessment will be the contrast between tirzepatide and placebo at Week 52. The MMRM analysis will be repeated using data during the on-treatment period, which is considered as up to the last dose date plus 7 days.

The MMRM model will include treatment, time (Weeks 24 and 52), treatment-by-time interaction, stratification factors as fixed effects, and baseline value of the KCCQ-CSS as a covariate. 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. An unstructured covariance structure will be used to model the within-patient errors. If the analysis fails to converge, the following variance-covariance matrices will be used until convergence is achieved:

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

In the MMRM analysis, the probability of missingness for any postbaseline KCCQ-CSS data that are not collected is assumed only to depend on the observed KCCQ-CSS values.

Change from baseline in the KCCQ-CSS will also be analyzed using an ANCOVA model. The ANCOVA model will include the categorical effect of treatment, stratification factors, and the continuous covariate of baseline KCCQ-CSS value. Missing KCCQ-CSS measurements at Week 52 will be imputed through multiple imputations as specified in Section 3.13.1.3.

3.13.1.5. Additional Analyses for Composite Endpoint of CV Death And HF Event

The contribution of each component of the primary composite endpoint (HF events and CV death) to the overall treatment effect will be examined. Methods similar to those described for the primary analysis will be used to separately analyze the time from randomization to the first occurrence of each component of the primary composite endpoint. The hazard ratio, with its CI and p-value, will be provided.

The CEC-confirmed total number of HF events (first and recurrent) and CV death will be analyzed by the semi-parametric proportional rates model (abbreviated as the LWYY model)

(Lin et al 2000) with treatment as a fixed effect adjusting for diagnosed T2DM (Y/N), baseline probability of HFpEF (<0.8, ≥0.8), and baseline NTproBNP (<200, ≥200 ng/L). A 30-day spacing rule will be applied when consider recurrent events. If an HF event is followed by another within 30 days, these HF events will be considered as a single HF event and the onset date will be the onset of the first HF event. If an HF event is followed by CV death within 30 days, only the CV death will be considered and the onset date will be the onset of the CV death. If an oral diuretic intensification is within 30 days of the date of randomization, it will not be considered as an endpoint. Participants who did not have an adjudicated primary endpoint event on or prior to end of follow-up will be censored at the date of participant's end of follow-up. The rate ratio with 95 confidence interval and p-value will be provided.

Non-parametric estimates of HF event rates over time, allowing for death as terminal event, will be provided (Ghosh and Lin 2000) with treatment as a fixed effect adjusting for diagnosed T2DM (Y/N), baseline probability of HFpEF (<0.8, ≥0.8), and baseline NTproBNP (<200, ≥200 ng/L). A similar spacing rule, as specified in Section 3.13.1.5, will be applied.

Analyses described in Section 3.13.1.2.1.2 for HF events and CV death will be repeated using the investigator's reported events.

3.13.2. Key Secondary Endpoints/Estimands

3.13.2.1. Change from Baseline in 6MWD at Week 52

3.13.2.1.1. Estimand for the Endpoints

The estimand is described by the following attributes:

- Endpoints: Change from baseline to Week 52 in 6MWD
- Treatment condition of interest: tirzepatide MTD vs. placebo
- Handling of intercurrent events: using treatment policy strategy to handle all intercurrent events
- Population: all randomized participants, and
- Population-level summary: median difference between treatment arms.

3.13.2.1.2. Main Analytical Approach

For the key secondary efficacy endpoints of change from baseline in 6MWD at Week 52, the same nonparametric approach as described in Section 3.13.1.1.2, and the same missing data imputation as described in Section 3.13.1.3, will be utilized.

3.13.2.2. Percent Change from Baseline in Body Weight Loss at Week 52

3.13.2.2.1. Estimand for the Endpoint

The estimand is described by the following attributes:

- Endpoint: percent change from baseline to Week 52 in body weight
- Treatment condition of interest: tirzepatide MTD versus. placebo
- Handling of ICEs: using treatment policy strategy to handle all ICEs
- Population: all randomized participants, and

- Population-level summary: the difference in means of the percent change between treatment arms.

3.13.2.2.2. Main Analytical Approach

The percent change from baseline in body weight will be analyzed using an analysis of covariance (ANCOVA) analysis. The ANCOVA model will include the categorical effect of treatment, stratification factors excluding baseline BMI group (<35, ≥35 kg/m^2), and the continuous covariate of baseline body weight value.

Missing body weight data at the scheduled postbaseline visits will be imputed using the retrieved dropout approach (as described in Section 3.13.1.3) through multiple imputation. The final statistical inference over multiple imputations will be obtained using the method proposed by Rubin (1987).

3.13.2.3. Change from Baseline in High-sensitivity C-reactive Protein at Week 52

3.13.2.3.1. Estimand for the Endpoint

- Endpoints: change from baseline to Week 52 in hsCRP.
- Treatment condition of interest: tirzepatide MTD versus placebo.
- Handling of ICEs: using the treatment policy strategy to handle all ICEs other than death.
- Population: all randomized participants.
- Population-level summary: the difference in mean change between treatment arms.

3.13.2.3.2. Main Analytical Approach

Change from baseline in hsCRP will be analyzed using an ANCOVA model. The ANCOVA model will include the categorical effect of treatment, stratification factors, and the continuous covariate of baseline hsCRP value. The ANCOVA model will be based on the log-transformed values of hsCRP.

Missing hsCRP at the scheduled postbaseline visits will be imputed using the retrieved dropout approach (as described in Section 3.13.1.3) through multiple imputation. The final statistical inference over multiple imputations will be obtained using the method proposed by Rubin (1987).

3.13.2.4. Sensitivity Analyses for Key Secondary Endpoints

For the change from baseline in 6MWD, a similar MMRM analysis as described in Section 3.13.1.4 will be conducted using data during the on-treatment period.

For percent change from baseline in body weight, a similar MMRM analysis as described in Section 3.13.1.4 will be conducted using data during the on-treatment period.

For the hsCRP change, a similar MMRM analysis as described in Section 3.13.1.4 will be conducted using data during the on-treatment period.

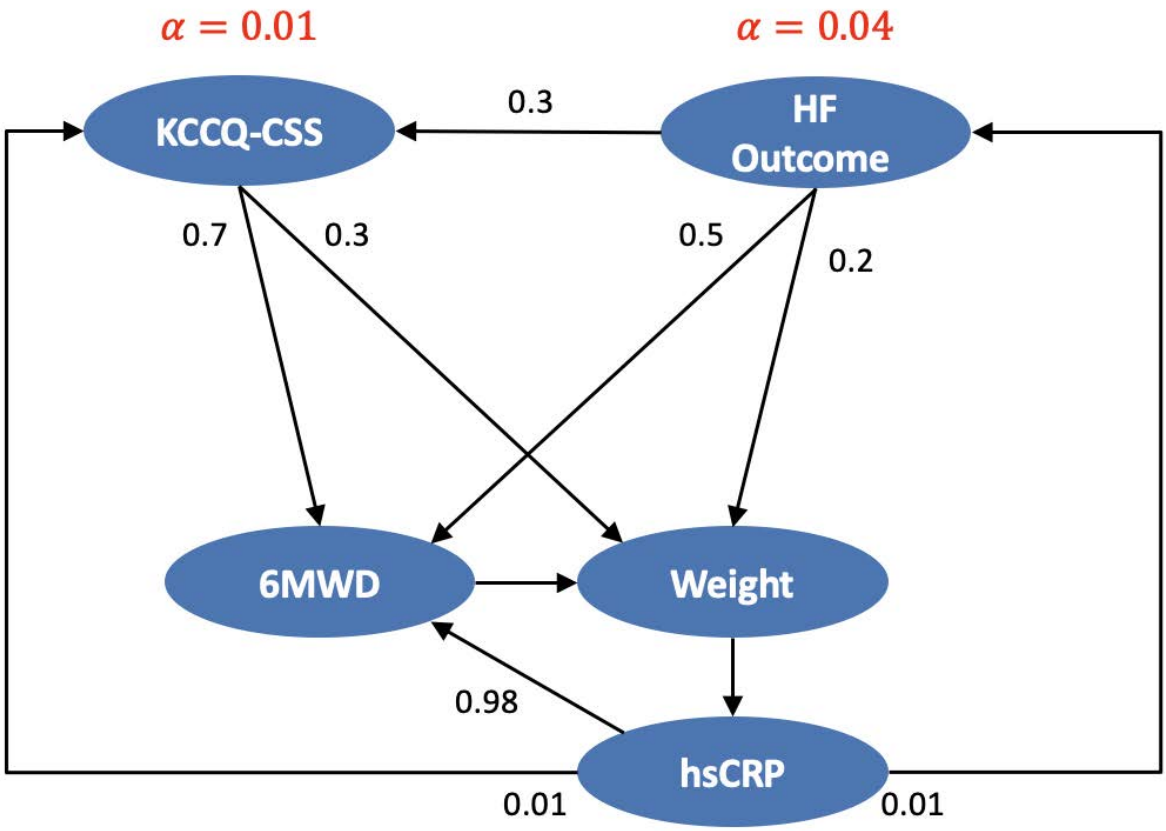
Change from baseline in the 6MWD will also be analyzed using an ANCOVA model. The ANCOVA model will include the categorical effect of treatment, stratification factors, and the

continuous covariate of the baseline 6MWD value. Missing 6MWD measurements at Week 52 will be imputed through multiple imputation as specified in Section 3.13.1.3.

3.13.3. Type I Error Rate Control Strategy for Primary and Key Secondary Efficacy Analyses

All primary and key secondary hypotheses will be tested with the overall family-wise type I error rate at a 2-sided alpha level of 0.05 through the multiplicity control approach based on the graphical multiple testing procedure. For the primary hypotheses, the HF outcome will be tested at a 2-sided alpha level of 0.04 and change in KCCQ-CSS will be tested at a 2-sided alpha level of 0.01 in parallel for statistical significance. If significant, the respective alpha of the primary endpoints will be propagated to test the key secondary endpoints. If any of the primary endpoints is not significant, then the appropriate alpha after the key secondary endpoints testing will be recycled to that primary endpoint.

Figure GPID.3.1 provides the details of the graphical multiple testing procedure.



Abbreviations: 6MWD = 6-minute walk distance; HF = heart failure ; hsCRP = high-sensitivity C-reactive protein; KCCQ-CSS = Kansas City Cardiomyopathy Questionnaire Clinical Summary Score.

Figure GPID.3.1 Graphical testing scheme for Study GPID.

3.13.4. Other Secondary

Unless otherwise specified, other secondary analyses will be guided by the treatment policy strategy using all randomized population. Missing data will not be imputed unless otherwise specified. The other secondary analyses can be seen in Table GPID.3.4.

Table GPID.3.4. Other Secondary

Objectives	Endpoints	Analytical Approaches
Hierarchical composite	A hierarchical composite of the following: • time to all-cause mortality through the end of follow-up • occurrence of HF events through end of follow-up, where HF events are as defined in Section 1.1 • number of HF events • time to first HF events • change from baseline in KCCQ-CSS category at Week 52, and • Change from baseline in the 6MWD category at Week 52. The categories for change from baseline in the KCCQ-CSS are: 1. ≥10-point worsening 2. ≥5- but <10-point worsening 3. No change (<5-point change) 4. ≥5- but <10-point improvement 5. ≥10- but <15-point improvement, and 6. ≥15-point improvement. The categories for change from baseline in the 6MWD are: 1. ≥30% worsening 2. ≥20% and <30% worsening 3. ≥10% and <20% worsening 4. no change (<10% change) 5. ≥10% and <20% improvement 6. ≥20% and <30% improvement, and 7. ≥30% improvement.	The win ratio (Pocock et al. 2012) will be reported as the measure of treatment effect based on the principle that each participant is compared with every other participant within each stratum in a pair-wise manner that proceeds in a hierarchical fashion. Participants will be stratified according to HF decompensation within 12 months of screening (Y/N), diagnosed T2DM (Y/N), and baseline BMI group (<35, ≥35 kg/m^2), yielding 8 stratification pools. The variance of win ratio will be calculated by the asymptotic normal U statistic approach (Dong et al. 2018). Missing KCCQ-CSS and 6MWD will be imputed as described in Section 3.13.1.3.
Clinical outcome events of HF	• Time to all-cause death • Time to first occurrence of HF events or all-cause death • Time to recurrent events of HF events and all-cause death • Time to first occurrence of HF events • Time to recurrent events of HF events	The time from randomization to the first occurrence of any component of the composite endpoint will be analyzed by a Cox proportional hazards model similar to the model described in Section 3.13.1.2.1.2. Time to recurrent event analyses will be performed using a LWYY as specified in Section 3.13.1.5. Similar 30-day spacing rule will be applied as described in Section 3.13.1.5.

Approved on 26 Jun 2024 GMT

Objectives	Endpoints	Analytical Approaches
NYHA Class	Proportion of participants with NYHA Class improvement at Week 52.	Logistic regression analysis will be conducted including treatment, stratification factors, and baseline NYHA Class in the model.
Weight loss	Proportion of participants attaining ≥5%, ≥10%, ≥15% and ≥20% in body weight reduction change at Week 52.	Logistic regression analysis will be conducted including treatment, stratification factors excluding baseline BMI group (<35, ≥35 kg/m^2), and the continuous covariate of baseline body weight value in the model. Missing body weight measurements at Week 52 will be imputed using a retrieved dropout approach (as described in Section 3.13.2.2) through multiple imputation.
Exercise capacity	• Change from baseline to Week 24 in 6MWD • Proportion of participants attaining 6MWD MWPC threshold at Week 52 • Proportion of participants attaining ≥10 meters, ≥20 meters, and ≥30 meters 6MWD change at Week 52	A similar nonparametric approach to the one described in Section 3.13.1.1.2 will be conducted. Logistic regression analysis will be conducted including treatment, stratification factors, and baseline 6MWD in the model. The MWPC threshold at Week 52 is decided in analysis conducted separately guided by the PAP. Missing 6MWD measurements at Week 52 will be imputed through multiple imputation, as specified in Section 3.13.1.3.
Patient-reported symptoms and physical limitations	• Change from baseline to Week 24 in KCCQ-CSS • Proportion of participants attaining KCCQ-CSS MWPC threshold at Week 52 • Proportion of participants attaining ≥5 meters, ≥10 meters, and ≥15 points KCCQ-CSS change at Week 52	A similar nonparametric approach to the one described in Section 3.13.1.1.2 will be conducted. Logistic regression analysis will be conducted including treatment, stratification factors, and baseline 6MWD in the model. The MWPC threshold at Week 52 is decided in analysis conducted separately guided by the PAP. Missing KCCQ-CSS measurements at Week 52 will be imputed through multiple imputation as specified in Section 3.13.1.3.

Abbreviations: 6MWD = 6-minute walk test distance; BMI = body mass index; HF = heart failure; KCCQ-CSS = Kansas City Cardiomyopathy Questionnaire – Clinical Summary Score; LWYY = Lin-Wei-Yang-Ying; MWPC = meaningful within-patient change; NYHA = New York Heart Association; PAP = psychometric analysis plan; T2DM = type 2 diabetes mellitus.

Note: MWPC threshold at Week 52 for KCCQ-CSS is 20; MWPC threshold at Week 52 for 6MWD is 25 meters.

3.13.5. Exploratory Endpoints

Unless otherwise specified, exploratory efficacy analyses will be guided by treatment policy strategy using all randomized population. Missing data will not be imputed.

Table GPID.3.5. Exploratory Efficacy Endpoints and Analyses

Objectives	Endpoints	Analytical Approaches
Atrial fibrillation	• Proportion of participants with atrial fibrillation at Week 24 and at Week 52. • Proportion of participants with atrial fibrillation at Week 24 and at Week 52 among those without atrial fibrillation at baseline. • Proportion of participants without atrial fibrillation at Week 24 and at Week 52 among those with atrial fibrillation at baseline.	Fisher's exact test or logistic regression analysis including treatment, stratification factors, and baseline atrial fibrillation status will be conducted.
Waist circumference	Change from baseline in waist circumference	MMRM analysis similar to the model described in Section 3.13.1.4 will be conducted.
Patient-reported health-related quality of life	Change from baseline in KCCQ: • Total Symptom Score (TSS) • Overall Summary Score (OSS)	MMRM analysis similar to the model described in Section 3.13.1.4 will be conducted.
Patient-reported health status	Change from baseline in EQ-5D-5L (measured at 24 weeks and 52 weeks): • Index Score, and • VAS Score.	MMRM analysis similar to the model described in Section 3.13.1.4 will be conducted.
PGIS	Proportion of participants with improvements from baseline in: • PGIS – Overall • PGIS – Physical Function, and • PGIS – Symptom Severity.	Proportion of participants with improvements from baseline will be summarized. Shift analysis from baseline to Week 24 and to Week 52 will also be performed.
Evaluation of prespecified biomarkers	Change from baseline in: • NT-proBNP • cTNT	MMRM analysis similar to the model described in Section 3.13.1.4 will be conducted. The data will be log-transformed for the analysis.
Waist to height ratio	Change from baseline to Week 52 in waist to height ratio	MMRM analysis similar to the model described in Section 3.13.1.4 will be conducted.
Kidney function	eGFR slope	The slope analysis will be constructed with eGFR as a dependent variable, including treatment group, stratification factors, baseline eGFR, time (as a continuous variable) and treatment-by-time interaction as fixed effects, and including subject as a random intercept and time as a random slope

Abbreviations: cTNT = cardiac troponin T; eGFR = estimated glomerular filtration rate; EQ-5D-5L = EuroQOL 5 Dimension 5 Level scale; KCCQ = Kansas City Cardiomyopathy Questionnaire; MMRM = mixed-effects model repeated measures; NT-proBNP = N-terminal pro-B-type natriuretic peptide; PGIS = Patient Global Impression of Severity; VAS = visual analog scale.
Note: Additional exploratory analysis may be conducted.

3.14. Safety Analyses

Unless specified otherwise, safety analyses will be conducted in the safety population (Table GPID.3.1). All events that occur between the first dose date of study drug and the end date of follow-up will be included in the safety analyses, regardless of the adherence to study drug.

The statistical assessment of homogeneity of the distribution of categorical safety responses between tirzepatide MTD and placebo will be conducted using Fisher's exact test, unless specified otherwise. Risk difference and its 95% CI will also be provided, where appropriate.

The mean change from baseline differences among treatments at all scheduled visits will be assessed via an MMRM analyses using REML. Unless specified otherwise, the MMRM analysis will contain measurements from planned visits up to 104 weeks, and the model will include treatment group, stratification factors, 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 3.13.1.4 will be tested in order until convergence has been met. If the data does not warrant the MMRM model, then an ANCOVA model will be conducted.

For selected safety parameters, a 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 follow-up.

3.14.1. Analysis of Adverse Events

3.14.1.1. Treatment-Emergent Adverse Events

A treatment-emergent AE (TEAE) is defined as an event that first occurred or worsened in severity after the first dose of study treatment. The 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. For events with a missing severity during the baseline period, the event 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 taking study medication for the first time, the case report form (CRF)-collected information (for example, treatment emergent flag, start time of study treatment, and event) will be used to determine whether the event was pre- versus posttreatment if available. If the relevant information is not available, then the events will be counted as posttreatment.

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

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

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

3.14.1.2. Common Adverse Events

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

3.14.1.3. Deaths

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

3.14.1.4. Other Serious Adverse Events

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

A listing of all SAEs will be provided. The listing will include treatment, participant identification including the site number, date of event, age at the time of enrollment, sex, AE group identification, MedDRA SOC and PT, severity, outcome, relationship to study drug, time from first dose of study drug to the event, and time from most recent dose to event (if the participant discontinued study drug prior to the event).

3.14.1.5. 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 the MedDRA PT nested within the SOC. Events will be ordered by decreasing frequency in the

tirzepatide arm within SOC. Additionally, a Kaplan-Meier plot of time to study treatment discontinuation due to AEs will be presented.

3.14.2. Patient Narratives

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

- death
- protocol defined serious adverse event,
- pregnancy, or
- permanent discontinuation of study treatment due to AEs.

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

3.14.3. Special Safety Topics

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

3.14.3.1. Amputation/Peripheral Revascularization

Amputation/peripheral revascularization will be considered as AESI. Participants with amputations/peripheral revascularization will be searched using the following MedDRA PTs and summarized:

- *Amputation*, and
- *Peripheral revascularization*.

A listing of participants with treatment-emergent (TE) events will be provided.

3.14.3.2. Diabetic Retinopathy Complications

Results of the baseline dilated fundoscopic exam will be summarized by treatment. Any TEAE suspected of worsening retinopathy triggers a follow-up dilated fundoscopic exam. A summary of TEAEs suspected of worsening retinopathy will be summarized by treatment and PT. Severe or serious AEs will be classified as AESI and a listing provided.

A complete list of PTs for inclusion in the above analyses is available in Appendix 2, Section 5.2.

3.14.3.3. Exocrine Pancreas Safety

3.14.3.3.1. Pancreatic Enzyme

Observed pancreatic enzyme data (p-amylase and lipase) will be summarized by treatment and nominal visit.

The counts and percentages of participants with maximum postbaseline pancreatic enzyme value exceeding the following thresholds will be provided by baseline pancreatic enzyme value (≤ upper limit of normal [ULN], > ULN), and postbaseline: ≤1 × ULN, (>1 to ≤3) × ULN, (> 3 to ≤5) × ULN, (>5 to ≤10) × ULN, >10 × ULN.

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

3.14.3.3.2. Pancreatitis Events

Investigator-reported events will be searched using the "*Acute pancreatitis*" Standardized MedDRA Query (SMQ) (20000022, narrow scope) and a "*Chronic pancreatitis*" PT in the AE database, while adjudication-confirmed pancreatitis are found from adjudication forms.

A summary of adjudicated pancreatic events will be provided by treatment.

TE adjudication confirmed pancreatitis will be considered as AESI. A listing of participants with investigator-reported and adjudicated pancreatitis will be provided.

3.14.3.4. Gastrointestinal Adverse Events

3.14.3.4.1. Nausea, Diarrhea, Constipation, and Vomiting

Summaries and analyses for incidence and severity of nausea, diarrhea (including "diarrhea", "diarrhoea" and "frequent bowel movements"), constipation (including "constipation", "infrequent bowel movement" and "faeces hard"), and vomiting (including "vomiting" and "vomiting projectile"), and nausea, vomiting and diarrhea combined, will be provided by each treatment group.

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

3.14.3.4.2. Severe Gastrointestinal Events

The PTs under the *Gastrointestinal (GI) disorders* SOC in MedDRA will be used to identify GI AEs, and only PTs with serious/severe TE cases will be considered as AESI.

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

3.14.3.5. Hepatobiliary Disorders

3.14.3.5.1. Hepatobiliary Events

The counts and percentages of participants with TE hepatic events will be summarized by treatment using the MedDRA PTs. The detailed search criteria can be found in Appendix 2, Section 5.2.

TE events related to acute gallbladder disease will also be summarized. The search criteria can be found in Appendix 2, Section 5.2.

Severe/serious TE hepatic events and severe/serious TE acute gallbladder disease will be considered as AESI and summarized separately.

In addition, counts and percentages of participants with acute gallbladder disease by weight change category will be provided by treatment.

3.14.3.5.2. Liver Enzymes

Common analyses for laboratory analyte measurements described in Section 3.14.1.5 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 study will be summarized between treatment groups:

- The counts and percentages of participants with an alanine transaminase (ALT) measurement ≥1× ULN, ≥3× ULN, ≥5× ULN, ≥10× ULN, and ≥20× ULN will be summarized for all participants with a postbaseline value and for subsets based on the following baseline values:
    - 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 (AST) measurement ≥1× ULN, ≥3× ULN, ≥5× ULN, ≥10× ULN, and ≥20 × ULN during the treatment period will be summarized for all participants with a postbaseline value and for subsets based on various baseline levels, as described above for ALT.
- The counts and percentages of participants with a total bilirubin (TBL) measurement ≥2× ULN, ≥5× ULN, and ≥8× ULN will be summarized for all participants with a postbaseline value and for the following subsets based on the 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 direct bilirubin (DBL) measurement ≥2× ULN AND ≥5× ULN will be summarized for all participants with a postbaseline value and for the same subsets as described for TBL.
- The counts and percentages of participants with a serum alkaline phosphatase (ALP) measurement ≥2× ULN and ≥3× ULN will be summarized for all participants with a postbaseline value and for the same subsets as described for TBL.
- The counts and percentages of participants with a gamma-glutamyltransferase (GGT) measurement ≥2× ULN will be summarized for all participants with a postbaseline value, if data available.

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

Hepatocellular drug-induced liver injury screening plot will be created. The plot with maximum postbaseline transaminase (ALT or AST, whichever is higher, regardless of the time between the two maximum values) divided by ULN versus maximum postbaseline total bilirubin divided by ULN (y-axis) will be created that includes all participants from the safety populations. Each participant with at least 1 postbaseline ALT or AST and total bilirubin contributes 1 point to the plot. Dashed lines represent TBL and transaminase cut-offs of 2× ULN and 3× ULN, respectively. A potential Hy's law case is circled and is defined as having a maximum postbaseline TBL equal to or exceeding 2× ULN within 30 days after maximum postbaseline ALT or AST equal to or exceeding 3× ULN, without cholestasis (defined as ALP less than 2× ULN). The percentages of study participants falling in each of the three relevant quadrants of the plot (right upper, left upper, right lower) will be summarized in a table.

Cholestatic Drug-Induced Liver Injury Screening plot based on the maximum postbaseline TBL and ALP will be created, regardless of the time between the two maximum values. Dashed lines represent TBL and ALP cut-offs of 2× ULN and 3× ULN, respectively. A potential cholestatic liver injury case is circled and is defined as having a maximum postbaseline TBL equal to or exceeding 2× ULN within 30 days after maximum postbaseline ALP equal to or exceeding 3× ULN. The percentages of study participants falling in each of the three relevant quadrants of the plot (right upper, left upper, right lower) will be summarized in a table.

3.14.3.6. Severe Persistent Hyperglycemia Requiring Rescue Therapy

A summary of initiation of rescue therapy in response to severe, persistent hyperglycemia will be provided by treatment group.

3.14.3.7. Hypoglycemia

The following categories in accordance with the 2021 American Diabetes Association position statement on glycemic targets (ADA 2021) will be defined in the database.

Glucose Alert Value (Level 1):

- **Documented symptomatic hypoglycemia** is defined as any time a participant feels 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) and ≥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 <70 mg/dL (<3.9 mmol/L) and ≥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 <70 mg/dL (<3.9 mmol/L) and ≥54 mg/dL (≥3.0 mmol/L).

Documented Clinically Significant Hypoglycemia (Level 2):

- **Documented symptomatic hypoglycemia** is defined as any time a participant feels 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. 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.

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

Hypoglycemia data will be censored at the time of receipt of new antihyperglycemic medication. This censoring will occur regardless of whether the reason for the alternative medication is rescue from severe, persistent hyperglycemia (with or without continuation of assigned study medication) or maintenance of glucose control in the event of cessation of assigned study medication.

Both the incidence (percent of participants experiencing ≥1 episode) and the rate (episodes/participant/year) of level 2 or level 3 hypoglycemia will be reported.

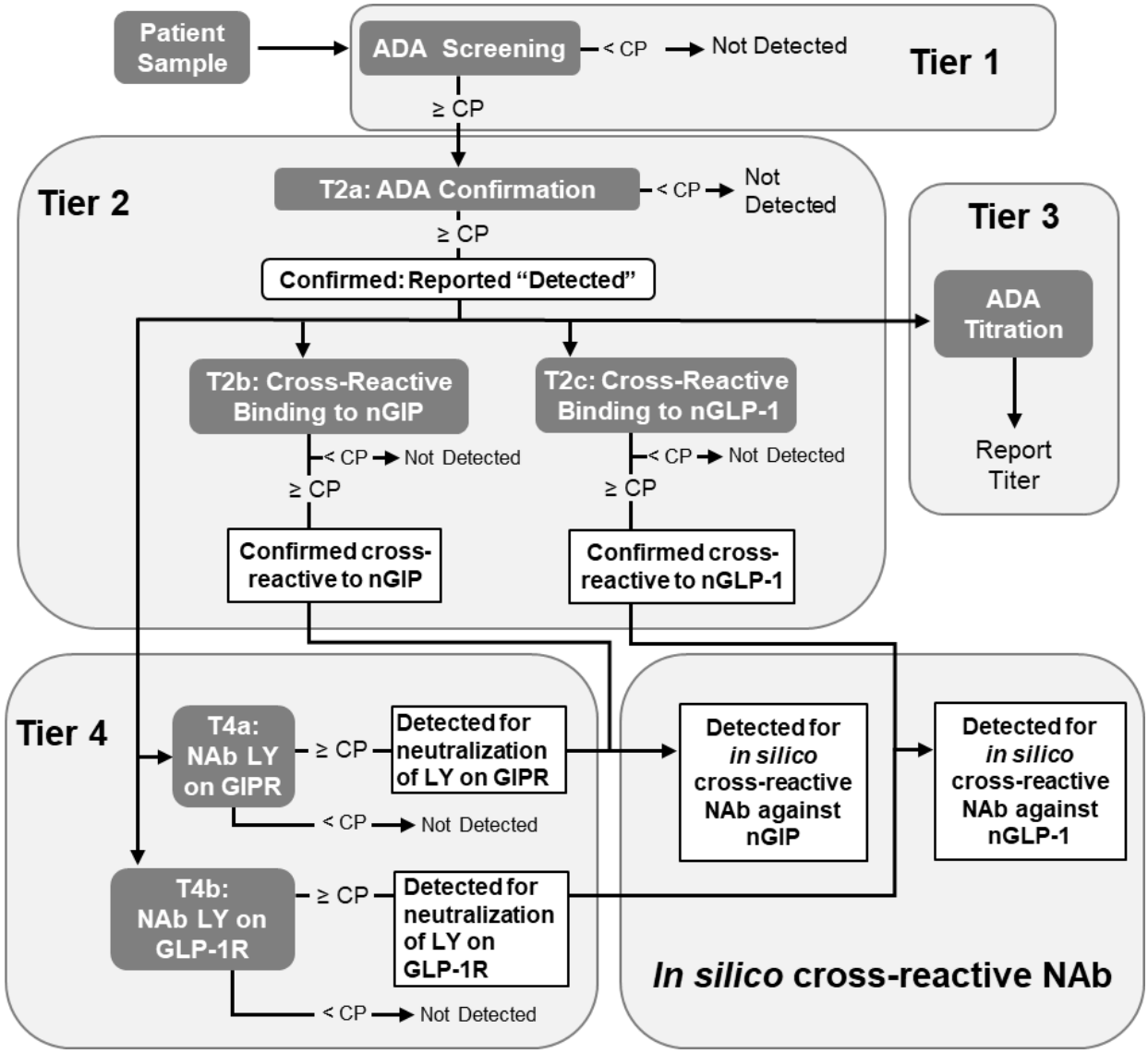
A listing of level 2 or 3 hypoglycemia will also be provided.

3.14.3.8. Immunogenicity

3.14.3.8.1. Definitions of Sample Anti-drug Antibody 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 and potentially multiple cross-reactive antibodies 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 GPID.3.2 details a flow chart that reflects the multitiered testing approach.



Abbreviations: ADA = antidrug antibodies; CP = cut point; GIPR = glucose-dependent insulinotropic polypeptide receptor; GLP-1R = glucagon-like peptide-1 receptor; LY = LY3298176 (tirzepatide); NAb = neutralizing antibodies; nGIP = native glucose-dependent insulinotropic polypeptide; nGLP-1 = native glucagon-like peptide-1; T2 = Tier 2; T4 = Tier 4.

Figure GPID.3.2. Flowchart of immunogenicity multitiered testing approach.

Table GPID.3.6 outlines results as reported from Tier 2a of the multitiered testing approach. Tier 4 results are reported similarly.

Table GPID.3.6. Sample 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 GPID.3.7).
NO TEST, QNS, etc.	Sample exists but was unevaluable by the assay.

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

It may be the case that the presence of high concentrations of tirzepatide will affect ADA immunoassays, and conversely, that high levels of ADA may affect the measurement of tirzepatide concentrations. Thus, a tirzepatide drug concentration, assessed from a sample drawn at the same time as the ADA sample, plays a key role in the clinical interpretation of a sample when the laboratory results is Not Detected (see Table GPID.3.7).

Table GPID.3.7. Sample Clinical ADA Interpretation Results

Sample Clinical Interpretation	Explanation
ADA Present	ADA assay result is Detected.
ADA Not Present	ADA assay result is Not Detected, and simultaneous drug concentration is at a level that has been demonstrated to not interfere in the ADA detection method (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 the 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 at a level that may cause interference in the ADA detection method.
ADA Missing	ADA sample not drawn, QNS, not tested, etc., causing there to be no laboratory result reported or the result is reported as "no test."

Abbreviations: ADA = antidrug antibodies; QNS = quantity not sufficient.

All ADA present samples will be evaluated for cross-reactive GIP (Tier 2b), cross-reactive GLP-1 (Tier 2c), NAb LY (tirzepatide) on GIP-R (Tier 4a), and NAb LY (tirzepatide) on GLP-1R (Tier 4b). If cross-reactive GIP is detected, NAb GIP on GIP-R (Tier 4c) is evaluated. If cross-reactive GLP-1 is detected, NAb GLP-1 on GLP-1R (Tier 4d) is evaluated (Figure GPID.3.2).

Similar terminology to Table GPID.3.8 applies for each type of cross-reactive and NAb assay. Importantly, each of these are distinct assays and, in general, have different assay operating characteristics.

The following are considered inconclusive for the NAb result:

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

To mitigate inconclusive cross-reactive NAb interpretations against native GIP and GLP-1 due to potential tirzepatide concentrations greater than or equal to the drug tolerance limit of the NAb GIP on GIP-R (Tier 4c) and NAb GLP-1 on GLP-1R (Tier 4d) assays, an *in silico* method utilizing results from Tiers 2b and 2c, Tiers 4a and 4b, and tirzepatide concentrations is introduced. The *in silico* method is outlined in Table GPID.3.8.

Table GPID.3.8. *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 NAb to nGIP	Tier 2b: "Not Detected"	Tier 4a "Not Detected" Or Tier 4a: "Detected" or N/A or Missing	Any Value or Missing	Not Present
	Tier 2b: "Detected"	Tier 4a: "Not Detected"	< drug tolerance limit of Tier 4a assay	Not Present
	Tier 2b: "Detected"	Tier 4a: "Not Detected"	≥ drug tolerance limit of Tier 4a assay	Inconclusive
	Tier 2b: "Detected"	Tier 4a: "Detected"	< drug tolerance limit of Tier 4a assay	Present
	Tier 2b: "Detected"	Tier 4a: "Detected"	≥ drug tolerance limit of Tier 4a assay	Present
Cross-Reactive NAb to nGLP-1	Tier 2c: "Not Detected"	Tier 4b "Not Detected" Or Tier 4b: "Detected" or N/A or Missing	Any Value or Missing	Not Present
	Tier 2c: "Detected"	Tier 4b: "Not Detected"	< drug tolerance limit of Tier 4b assay	Not Present
	Tier 2c: "Detected"	Tier 4b: "Not Detected"	≥ drug tolerance limit of Tier 4b assay	Inconclusive
	Tier 2c: "Detected"	Tier 4b: "Detected"	< drug tolerance limit of Tier 4b assay	Present
	Tier 2c: "Detected"	Tier 4b: "Detected"	≥ drug tolerance limit of Tier 4b assay	Present

Abbreviations: ADA = antidrug antibodies; GIPR = glucose-dependent insulinotropic polypeptide receptor; GLP-1R = glucagon-like peptide-1 receptor; LY = tirzepatide; 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: Only the drug tolerance limits of the Tier 4a and 4b assays are used for in silico classifications as they are lower than the drug tolerance limits of the Tier 2b and 2c assays, respectively.

3.14.3.8.2. 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, the last nonmissing immunogenicity assessment prior to first administration of study drug is used to determine participant's ADA status and TE status (see below).

Immunogenicity Postbaseline Period Observations: Postbaseline period observations for each participant includes all observations after the first administration of study drug.

3.14.3.8.3. Definitions of Participant ADA Status

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

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

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

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

TE ADA positive (TE ADA+) participant: an evaluable participant who had a:

- baseline status of ADA Not Present and at least 1 postbaseline status of ADA Present with titer ≥2× minimum required dilution (MRD), where the MRD is the minimum required dilution of the ADA assay, or
- 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 (B) status of ADA Present, with titer 1:B, and at least 1 postbaseline (P) status of ADA Present, with titer 1:P, with P/B ≥4.

As shown in Figure GPID.3.2, a titer is expected when the ADA assay result is Detected. On the occasion when the corresponding assay cannot be performed, in which case a titer value will be imputed for the purpose of TE ADA determination. A baseline sample with detected ADA and no titer is imputed to be the MRD (1:10), and a postbaseline sample with ADA detected and no titer is imputed to be one dilution above the MRD (1:20).

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

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

For each NAb assay, the following are defined:

NAb positive (NAb+) participant: a participant who is TE ADA+ and has a 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 a NAb Inconclusive sample result.

NAb negative (NAb-) participant: a participant is neither NAb+ nor NAb Inconclusive.

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

3.14.3.8.4. Analyses to be Performed

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

The cross-reactive NAb will exclude Tier 4c and 4d results but include the *in silico* classification as cross-reactive NAb for summary.

A summary will be provided of the count and percentage of tirzepatide-treated participants experiencing specific TEAE (see Table GPID.3.9) by participant TE ADA status (TE ADA+, TE ADA-, TE ADA Inconclusive). The PT will be ordered by decreasing incidence in TE ADA+ status group.

Table GPID.3.9. Adverse Events for Analysis with Immunogenicity Results

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

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

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

3.14.3.9. Hypersensitivity Reactions

Two main analyses are performed in support of the 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 the start of study drug administration up to 24 hours after the end of study drug administration. For events without the hypersensitivity eCRF, only date (no time) information is collected. Among these events without time information, if the event occurred on the same date as the study drug injection date it 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.

Analyses for both time periods are based on the following:

- narrow and algorithm terms in *Anaphylactic reaction* SMQ (20000021) (analysis for algorithm term only applicable for Time Period A)
- narrow terms in *Angioedema* SMQ (20000024)
- narrow terms in *Severe cutaneous adverse reactions* SMQ (20000020)
- narrow terms in *Hypersensitivity* SMQ (20000214), and
- narrow terms in *Vasculitis* SMQ (20000174).

For the *Anaphylactic reaction* SMQ, each term is classified by scope (narrow, broad) and by category (A, B, C, D). All narrow terms are category A, and all broad terms are category B, C, or D. In addition to the usual narrow and broad searches, the SMQ defines an algorithm to further refine the cases of interest. For the 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 4 SMQs indicated above (that is, combined search across narrow of all 4 SMQs)
- any narrow scope term within each SMQ, separately (that is, narrow SMQ search). For the Time Period A analysis, any term from Anaphylactic reaction SMQ algorithm.

Within each query, individual PTs that satisfied the queries will be summarized. For the 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.

3.14.3.9.1. Severe/Serious Hypersensitivity Reactions

The severe or serious cases of hypersensitivity will be considered as AESI. Summaries of severe/serious TE hypersensitivity reactions by treatment will be provided.

3.14.3.10. Injection Site Reactions

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 location of the reaction, timing of the reaction relative to study drug administration, and characteristics of the injection site reaction: erythema, induration, pain, pruritus, and edema.

Patient-based analysis and event-based analysis may be provided if necessary. The patient-based analysis summarizes all injection-site reaction (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 the differential contribution of individual patients to the analysis.

The counts and percentages of participants with TE ISRs will be summarized by treatment using MedDRA PTs. Detailed search criteria can be found in Appendix 2, Section 5.2.

Events will be ordered by decreasing frequency in the tirzepatide arm.

3.14.3.10.1. Severe/Serious Injection Site Reactions

The severe or serious ISRs (for example, abscess, cellulitis, erythema, hematomas/hemorrhage, exfoliation/necrosis, pain, subcutaneous nodules, swelling, induration, inflammation) will be considered as AESI.

The counts and percentage of participants with severe or serious TE ISRs will be summarized by treatment.

3.14.3.11. Major Adverse Cardiovascular Events

Major adverse cardiovascular events (MACE) reported by investigators are adjudicated by an independent CEC in a blinded fashion.

The following positively-adjudicated MACE will be considered as an AESI:

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

Cardiovascular death and hospitalization for HF are not considered as AESI since they are reported as efficacy endpoints in Study GPID.

The counts and percentages of participants with adjudicated MACE may be summarized by treatment. In addition, MACE reported by investigator may also be summarized, although MACE reported by investigator is not considered as AESI.

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

3.14.3.12. Major Depressive Disorder/Suicidal Ideation or Behavior

The severe or serious TE major depressive disorder/suicidal ideation or behavior will be captured as AESI. AEs will be searched using MedDRA PTs from SMQs narrow scope: 20000037 (*Suicide/self-injury*) and 20000167 (*Depression [excl suicide and self injury]*).

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 in the tirzepatide arm nested within SMQ.

3.14.3.13. Renal Safety

Laboratory measures related to renal safety will be analyzed as specified for laboratory measurements in Section 3.14.5.

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

MMRM analyses as described in Section 3.14 for eGFR and UACR will be provided. Log-transformation will be performed for UACR.

3.14.3.13.1. Acute Renal Events

As severe GI events may lead to dehydration, which could cause a deterioration in renal function including acute renal failure, dehydration events will be analyzed. Acute renal events associated with chronic renal failure exacerbation will also be captured.

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

- Acute renal failure: narrow terms in *Acute renal failure* SMQ (20000003), and
- Chronic kidney disease: narrow terms in *Chronic kidney disease* SMQ (20000213).

Severe or serious renal events will be considered as AESI and will also be summarized by treatment

3.14.3.13.2. Dehydration

Dehydration events will be captured in the narrow terms in *Dehydration* SMQ (20000232).

The counts and percentages of participants with dehydration events will be summarized by treatment. Severe or serious TE dehydration will be considered as AESIs and summarized separately.

3.14.3.14. Thyroid Safety Monitoring

3.14.3.14.1. Calcitonin

The purpose of calcitonin measurements is to assess the potential of tirzepatide to affect thyroid C-cell function, which may indicate development of C-cell hyperplasia and neoplasms.

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

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

3.14.3.14.2. C-Cell Hyperplasia and Thyroid Malignancies

TE thyroid malignancies and C-cell hyperplasia will be considered as AESI. Thyroid malignancies and C-cell hyperplasia will be identified using the MedDRA high-level term (HLT) for *Thyroid neoplasms malignant* and PT for *Thyroid C-cell hyperplasia*.

The counts and percentages of participants with TE thyroid C-cell hyperplasia and malignancies will be summarized.

3.14.3.15. Arrhythmias and Cardiac Conduction Disorders

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

The TE arrhythmias and cardiac conduction disorders events will be identified using the MedDRA PTs. Detailed searching criteria can be found in Appendix 2, Section 5.2.

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 PTs will be ordered with decreasing frequency in the tirzepatide arm within SMQ.

3.14.3.16. Malignancy

The malignancy events will be included using the MedDRA PTs contained in the *Malignant tumours* SMQ (20000194) narrow scope or *Tumours of unspecified malignancy* SMQ (20000195) narrow scope. Malignancy will be considered as an AESI.

The counts and percentages of participants with TE malignancy will be summarized by treatment.

3.14.3.17. Metabolic Acidosis

Severe/serious metabolic acidosis, including diabetic ketoacidosis, will be captured as AESI.

A listing of participants with TE events will be provided using the list of PTs provided in Appendix 2, Section 5.2.

3.14.3.18. Hypotension, Orthostatic Hypotension, and Syncope

The AE database will be searched using predefined PTs to identify events consistent with hypotension, orthostatic hypotension, and syncope.

The counts and percentages of participants with TE hypotension, orthostatic hypotension, and syncope will be summarized by treatment and PT. Detailed searching criteria can be found in Appendix 2, Section 5.2.

3.14.4. Vital Signs

Two sitting blood pressure and apical heart rate measurements are collected at each visit scheduled for vital sign collection. For the multiple records of an individual vital sign collected at the same visit, the average value will be used for data summaries and analyses. The vital signs collected in association with 6-minute walk test (6MWT) will be excluded from the vital signs analyses.

Descriptive summaries by treatment and by nominal scheduled visit will be provided for baseline and postbaseline values as well as change from baseline values.

An MMRM and/or an ANCOVA model as described in Section 3.14 may be conducted if necessary.

Counts and percentages of participants with abnormal sitting systolic blood pressure, sitting diastolic blood pressure, and heart rate will be presented by treatment for participants who have both baseline and at least 1 postbaseline result at any time during the entire study. To assess decreases, change from the minimum value during the baseline period to the minimum value during the postbaseline period will be used. To assess increases, changes from the maximum value during the baseline period to the maximum value during the postbaseline period will be used. Both planned and unplanned measurements will be included in the analysis. The criteria for identifying participants with vital signs abnormalities are listed in Table GPID.3.10.

Table GPID.3.10. Categorical Criteria for Abnormal Blood Pressure and Heart Rate 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
Heart rate (bpm) (apical)	<50 and decrease from baseline ≥15	>100 and increase from baseline ≥15

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

Additional analyses of heart rate

Counts and percentages of participants who have changes from baseline to postbaseline or postbaseline absolute heart rate will be summarized by treatment group for the categories listed in Table GPID.3.11.

Table GPID.3.11. Categorical Criteria for Additional Analyses of Heart Rate

Value at any postbaseline visit	>100 bpm, >130 bpm
Value at 2 consecutive postbaseline visits	>100 bpm
Value at any 3 postbaseline visits	>100 bpm
Change from baseline to maximum at any visit	>20 bpm
Change from baseline at 2 consecutive visits	>20 bpm
Change from baseline at any 3 visits	>20 bpm

Abbreviation: bpm = beats per minute.

3.14.5. Clinical Laboratory Evaluation

All safety laboratory data will be reported using the International System of Units and conventional units. Limits from the performing lab will be used to define low (L) and high (H). Descriptive summaries by treatment and by nominal visit will be provided for the baseline and postbaseline values as well as the change from baseline values.

For selected laboratory analyte measurements collected quantitatively, observed and change from baseline values for each visit may be displayed in plots for participants who have both a baseline and at least 1 postbaseline planned measurement. Baseline will be defined as the last nonmissing observation during the baseline period. 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 with a change from baseline to postbaseline of normal or high to low, and normal or low to high. The proportion of participants shifted will be compared between treatments using Fisher's exact test.

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

The MMRM model or ANCOVA (if the MMRM model is not applicable), as described in Section 3.14, will be used for the analysis for the continuous measurements for selected lab tests with or without log-transformed (postbaseline measure/baseline measure) response variables. For measures analyzed using log-transformed values, the results will be presented with the scale back-transforming to the original, related scale.

3.15. Subgroup Analyses

3.15.1. Primary Efficacy Endpoints

The following subgroup variables will be considered for subgroup analyses if there are adequate number of participants in each subcategory:

- age group: <65, ≥65
- race: White, American Indian or Alaska Native, Black or African American, Asian, or Other
- sex: Male, Female
- ethnicity: Hispanic or Latino, Not Hispanic or Latino
- region: US, Central/South America, Asia, and Other
- baseline BMI (kg/m^2): <35, ≥35
- baseline BMI (kg/m^2): <35, ≥35 and <40, ≥40
- HF decompensation within 12 months of screening: Yes, No
- diagnosed T2DM at screening: Yes, No
- atrial fibrillation at baseline: Yes, No
- baseline eGFR (mL/min/1.73m^2): <60, ≥60
- N-terminal pro-B-type natriuretic peptide (NT-proBNP) at baseline: <200, ≥200
- NYHA class: Class II, Class III or IV
- baseline use of MRA: Yes, No
- baseline use of RAS inhibitors (ACE +ARB+ARNi): Yes, No
- baseline use of a beta blocker: Yes, No
- baseline use of sodium-glucose co-transporter 2 inhibitors: Yes, No
- baseline use of diuretics: Yes, No
- baseline left ventricular ejection fraction: <60, ≥60
- baseline systolic blood pressure: <130, ≥130
- baseline hsCRP: <2, ≥2
- baseline KCCQ-CSS: <Median, ≥Median
- coronary artery disease at baseline: Yes, No
- baseline heart rate: <70, ≥70
- baseline probability of HFpEF: <0.8, ≥0.8, and
- baseline waist-to-height ratio: <0.6, ≥0.6.

Subgroup analyses will be considered for the primary efficacy endpoints. An ANCOVA analysis similar to the model described in Section 3.13.1.4 will be considered for the KCCQ-CSS. A Cox proportional hazards model similar to the model described in Section 3.13.1.2 will be considered for the HF outcomes in each subcategory of the subgroup variable with more than 10 events for the corresponding primary endpoint. In addition, a full model with additional terms of subgroup and subgroup-by-treatment interactions will be used to obtain interaction p-values.

The subgroup analyses may also be performed for the key secondary efficacy endpoints using the primary analysis approaches defined in Section 3.13.2. The same analysis model for the corresponding endpoint will be conducted in each subcategory of the subgroup variable to obtain

estimates of the treatment group difference. In addition, a full model with additional terms of subgroup and subgroup-by-treatment interactions will be used to obtain interaction p-values.

3.15.2. Safety in Special Groups and Situations

3.15.2.1. Intrinsic Factors

A subgroup analysis will be presented for common TEAEs, for the purposes of presentation in the summary of clinical safety (SCS). The subgroups will be

- age group: <65, ≥65 years
- sex: female, male
- race: American Indian or Alaska Native, Asian, Black/African American, Native Hawaiian or Other Pacific Islander, White, and Other
- BMI: ≥30 and <35, ≥35 and <40, ≥40 kg/m^2
- T2DM: Yes, No, and
- eGFR: <60 mL/min/1.73m^2, ≥60 mL/min/1.73m^2.

The response variable will be each most common TEAE. The explanatory variables will be treatment, subgroup, and treatment-by-subgroup interaction. Within each subgroup category, odds ratios (treatment over placebo) and associated p-values will be provided.

An additional summary of AEs will be provided for the age groups of <65, ≥65 to <75, ≥75 to <85, and ≥85 years to meet expectations from the European Union (EMA 2014). A summary table will be created similarly to Table 12.12 in the PHUSE AE white paper (PHUSE 2017). Rows of the table include number of participants with at least 1 of the following:

- TEAE
- SAE (separate rows for fatal, hospitalization, life-threatening, disability, other)
- AEs leading to study intervention discontinuation
- Accidents and injuries (SMQ)
- Cardiac disorders (SOC)
- Infections and infestations (SOC)
- Nervous system disorders (SOC)
- Psychiatric disorders (SOC)
- Vascular disorders (SOC)
- Central nervous system vascular disorders (SMQ)
- Anticholinergic syndrome (PT),
- Fractures, or
- Hypotension, falls, fractures.

3.15.2.2. Extrinsic Factors

Subset analyses will be conducted for region and ethnicity for the SCS.

The following regions will be considered:

- United States
- Central/South America (Argentina, Brazil, Mexico)
- Asia (China, India, Taiwan), and
- Other (Israel, Russia).

3.16. Interim Analyses and Data Monitoring Committee

The details for the interim analyses and data monitoring committee (DMC) will be provided in the DMC charter.

4. Unblinding Plan

Details of the blinding and unblinding will be provided in the blinding and unblinding plan document for Study GPID.

5. Supporting Documentation

5.1. Appendix 1: Kansas City Cardiomyopathy Questionnaire - Scoring Instructions

The Kansas City Cardiomyopathy Questionnaire
Scoring Instructions

There are 10 summary scores within the KCCQ, which are calculated as follows:

1. **Physical Limitation**

   - Code responses to each of Questions 1a-f as follows:

     Extremely limited = 1
     Quite a bit limited = 2
     Moderately limited = 3
     Slightly limited = 4
     Not at all limited = 5
     Limited for other reasons or did not do = <*missing value*>

   - If at least three of Questions 1a-f are not missing, then compute

     Physical Limitation Score = 100*[(mean of Questions 1a-f actually answered) – 1]/4

     (see footnote at end of this document for explanation of meaning of "actually answered")

2. **Symptom Stability**

   - Code the response to Question 2 as follows:

     Much worse = 1
     Slightly worse = 2
     Not changed = 3
     Slightly better = 4
     Much better = 5
     I've had no symptoms over the last 2 weeks = 3

   - If Question 2 is not missing, then compute

     Symptom Stability Score = 100*[(Question 2) – 1]/4

3. **Symptom Frequency**

   - Code responses to Questions 3, 5, 7 and 9 as follows:

     <u>Question 3</u>
     Every morning = 1
     3 or more times a week but not every day = 2
     1-2 times a week = 3
     Less than once a week = 4
     Never over the past 2 weeks = 5

3. Symptom Frequency (cont.)

Questions 5 and 7
All of the time = 1
Several times a day = 2
At least once a day = 3
3 or more times a week but not every day = 4
1-2 times a week = 5
Less than once a week = 6
Never over the past 2 weeks = 7

Question 9
Every night = 1
3 or more times a week but not every day = 2
1-2 times a week = 3
Less than once a week = 4
Never over the past 2 weeks = 5

- If at least two of Questions 3, 5, 7 and 9 are not missing, then compute:

  S3 = [(Question 3) – 1]/4
  S5 = [(Question 5) – 1]/6
  S7 = [(Question 7) – 1]/6
  S9 = [(Question 9) – 1]/4

  Symptom Frequency Score = 100*(mean of S3, S5, S7 and S9)

4. Symptom Burden

- Code responses to each of Questions 4, 6 and 8 as follows:

  Extremely bothersome = 1
  Quite a bit bothersome = 2
  Moderately bothersome = 3
  Slightly bothersome = 4
  Not at all bothersome = 5
  I've had no swelling/fatigue/shortness of breath = 5

- If at least one of Questions 4, 6 and 8 is not missing, then compute

  Symptom Burden Score = 100*[(mean of Questions 4, 6 and 8 actually answered) – 1]/4

5. Total Symptom Score

= mean of the following available summary scores:
Symptom Frequency Score
Symptom Burden Score

 Rev. 3/27/01

6. Self-Efficacy

- Code responses to Questions 10 and 11 as follows:

  Question 10
  Not at all sure = 1
  Not very sure = 2
  Somewhat sure = 3
  Mostly sure = 4
  Completely sure = 5

  Question 11
  Do not understand at all = 1
  Do not understand very well = 2
  Somewhat understand = 3
  Mostly understand = 4
  Completely understand = 5

- If at least one of Questions 10 and 11 is not missing, then compute

  Self-Efficacy Score = 100*[(mean of Questions 10 and 11 actually answered) – 1]/4

7. Quality of Life

- Code responses to Questions 12, 13 and 14 as follows:

  Question 12
  It has extremely limited my enjoyment of life = 1
  It has limited my enjoyment of life quite a bit = 2
  It has moderately limited my enjoyment of life = 3
  It has slightly limited my enjoyment of life = 4
  It has not limited my enjoyment of life at all = 5

  Question 13
  Not at all satisfied = 1
  Mostly dissatisfied = 2
  Somewhat satisfied = 3
  Mostly satisfied = 4
  Completely satisfied = 5

  Question 14
  I felt that way all of the time = 1
  I felt that way most of the time = 2
  I occasionally felt that way = 3
  I rarely felt that way = 4
  I never felt that way = 5

 Rev. 3/27/01

7. Quality of Life (cont.)

- If at least one of Questions 12, 13 and 14 is not missing, then compute

  Quality of Life Score = 100*[(mean of Questions 12, 13 and 14 actually answered) – 1]/4

8. Social Limitation

- Code responses to each of Questions 15a-d as follows:

  Severely limited = 1
  Limited quite a bit = 2
  Moderately limited = 3
  Slightly limited = 4
  Did not limit at all = 5
  Does not apply or did not do for other reasons = <*missing value*>

- If at least two of Questions 15a-d are not missing, then compute

  Social Limitation Score = 100*[(mean of Questions 15a-d actually answered) – 1]/4

9. Overall Summary Score

= mean of the following available summary scores:
Physical Limitation Score
Total Symptom Score
Quality of Life Score
Social Limitation Score

10. Clinical Summary Score

= mean of the following available summary scores:
Physical Limitation Score
Total Symptom Score

Note: references to "**means of questions actually answered**" imply the following.

- If there are *n* questions in a scale, and the subject must answer *m* to score the scale, but the subject answers only *n-i*, where *n-i* >= *m*, calculate the **mean of those questions** as
  (sum of the responses to those *n-i* questions) / (*n-i*)
  not
  (sum of the responses to those *n-i* questions) / *n*

If doing these calculations seems like too much trouble, consider using one of our tools – available at www.cvoutcomes.org:

- SAS or SPSS code
- Excel spreadsheets
- Web data services

 Rev. 3/27/01

Question No.	Domain	Question	Response
1a	Physical limitation	How much are you limited by heart failure (shortness of breath or fatigue) in dressing yourself over the past 2 weeks?	1: Extremely limited 2: Quite a bit limited 3: Moderately limited 4: Slightly limited 5: Not at all limited 0: Limited for other reasons or did not do the activity
1b	Physical limitation	How much are you limited by heart failure (shortness of breath or fatigue) in showering/bathing over the past 2 weeks?	
1c	Physical limitation	How much are you limited by heart failure (shortness of breath or fatigue) in walking 1 block on level ground over the past 2 weeks?	
1d	Physical limitation	How much are you limited by heart failure (shortness of breath or fatigue) in doing yard work, housework or carrying groceries over the past 2 weeks?	
1e	Physical limitation	How much are you limited by heart failure (shortness of breath or fatigue) in climbing a flight of stairs without stopping over the past 2 weeks?	
1f	Physical limitation	How much are you limited by heart failure (shortness of breath or fatigue) in hurrying or jogging over the past 2 weeks?	

Question No.	Domain	Question	Response
2	Symptom Stability	Compared with 2 weeks ago, have your symptoms of heart failure (shortness of breath, fatigue, or ankle swelling) changed?	1: Much worse 2: Slightly worse 3: Not changed 4: Slightly better 5: Much better 3.0: I had no symptoms over the last 2 weeks
3	Symptom Frequency	Over the past 2 weeks, how many times did you have swelling in your feet, ankles or legs when you woke up in the morning?	1: Every morning 2: 3 or more times a week, but not every day 3: 1-2 times a week 4: Less than once a week 5: Never over the past 2 weeks
4	Symptom Burden	Over the past 2 weeks, how much has swelling in your feet, ankles or legs bothered you?	1: Extremely bothersome 2: Quite a bit bothersome 3: Moderately bothersome 4: Slightly bothersome 5: Not at all bothersome 5.0: I've had no swelling

Question No.	Domain	Question	Response
5	Symptom Frequency	Over the past 2 weeks, on average, how many times has fatigue limited your ability to do what you want?	1: All of the time, 2: Several times per day 3: At least once a day 4: 3 or more times per week but not every day 5: 1-2 times per week 6: Less than once a week 7: Never over the past 2 weeks
6	Symptom Burden	Over the past 2 weeks, how much has your fatigue bothered you?	1: Extremely bothersome 2: Quite a bit bothersome 3: Moderately bothersome 4: Slightly bothersome 5: Not at all bothersome 5.0: I've had no fatigue
7	Symptom Frequency	Over the past 2 weeks, on average, how many times has shortness of breath limited your ability to do what you wanted?	1: All of the time 2: Several times per day 3: At least once a day 4: 3 or more times per week but not every day 5: 1-2 times per week 6: Less than once a week 7: Never over the past 2 weeks

Question No.	Domain	Question	Response
8	Symptom Burden	Over the past 2 weeks, how much has your shortness of breath bothered you?	1: Extremely bothersome 2: Quite a bit bothersome 3: Moderately bothersome 4: Slightly bothersome 5: Not at all bothersome 5.0: I've had no shortness of breath
9	Symptom Frequency	Over the past 2 weeks, on average, how many times have you been forced to sleep sitting up in a chair or with at least 3 pillows to prop you up because of shortness of breath?	1: Every night 2: 3 or more times a week, but not every day 3: 1-2 times a week 4: Less than once a week 5: Never over the past 2 weeks
10	Self-efficacy	Heart failure symptoms can worsen for a number of reasons. How sure are you that you know what to do, or whom to call, if your heart failure gets worse?	1: Not at all sure 2: Not very sure 3: Somewhat sure 4: Mostly sure 5: Completely sure
11	Self-efficacy	How well do you understand what things you are able to do to keep your heart failure symptoms from getting worse?	1: Do not understand at all 2: Do not understand very well 3: Somewhat understand 4: Mostly understand 5: Completely understand

Question No.	Domain	Question	Response
12	Quality of Life	Over the past 2 weeks, how much has your heart failure limited your enjoyment of life?	1: It has extremely limited my enjoyment of life 2: It has limited my enjoyment of life quite a bit 3: It has moderately limited my enjoyment of life 4: It has slightly limited my enjoyment of life 5: It has not limited my enjoyment of life at all
13	Quality of Life	If you had to spend the rest of your life with your heart failure the way it is right now, how would you feel about this?	1: Not at all satisfied 2: Mostly dissatisfied 3: Somewhat satisfied 4: Mostly satisfied 5: Completely satisfied
14	Quality of Life	Over the past 2 weeks, how often have you felt discouraged or down in the dumps because of your heart failure?	1: I felt that way all of the time 2: I felt that way most of the time 3: I occasionally felt that way 4: I rarely felt that way 5: I never felt that way

Question No.	Domain	Question	Response
15a	Social Limitation	How much have your heart failure limited your participation in hobbies, recreational activities over the past 2 weeks?	1: Severely limited 2: Limited quite a bit 3: Moderately limited 4: Slightly limited 5: Did not limited at all 6: Does not apply or did not do for other reasons
15b	Social Limitation	How much have your heart failure limited your participation in working or doing household chores over the past 2 weeks?	
15c	Social Limitation	How much have your heart failure limited your participation in visiting family or friends out of your home over the past 2 weeks?	
15d	Social Limitation	How much have your heart failure limited your participation in intimate relationships with loved ones over the past 2 weeks?	

5.2. Appendix 2: Search Criteria for Special Safety Topics

Diabetic Retinopathy Complications

Diabetic retinopathy complications will be identified using the following MedDRA PTs:

Preferred Term	Code
Arteriosclerotic retinopathy	10063452
Blindness	10005169
Blindness transient	10005184
Blindness unilateral	10005186
Cystoid macular oedema	10058202
Diabetic blindness	10012646
Diabetic eye disease	10012661
Diabetic retinal oedema	10012688
Diabetic retinopathy	10012689
Diabetic uveitis	10012692
Exudative retinopathy	10015901
Eye laser surgery	10057105
Fundoscopy	10017519
Fundoscopy abnormal	10017520
Intra-ocular injection	10057098
Macular detachment	10075873
Macular oedema	10025415
Maculopathy	10025425
Non-proliferative retinopathy	10081568
Phacotrabeculectomy	10059276
Retinal collateral vessels	10077911
Retinal cryoablation	10074908
Retinal detachment	10038848
Retinal exudates	10038862
Retinal haemorrhage	10038867
Retinal laser coagulation	10038873
Retinal neovascularisation	10055666
Retinal oedema	10038886
Retinal operation	10062107
Retinal thickening	10077890
Retinal vascular disorder	10038901
Retinal vascular occlusion	10038903
Retinal vein occlusion	10038907
Retinopathy	10038923

Preferred Term	Code
Retinopathy haemorrhagic	10051447
Retinopathy hypertensive	10038926
Retinopathy proliferative	10038934
Sudden visual loss	10042441
Vision blurred	10047513
Visual acuity reduced	10047531
Visual acuity reduced transiently	10047532
Visual impairment	10047571
Vitrectomy	10047644

Arrhythmias and Cardiac Conduction Disorders

TE arrhythmias, arrhythmias, and cardiac conduction disorders will be considered as AESI. Cardiovascular events will include clinically relevant rhythm and conduction disorders. TE arrhythmias and cardiac conduction disorders events will be included using the MedDRA PT contained in any of the following SMQs:

1) Arrhythmias:
   - For symptoms: *Arrhythmia related investigations, signs and symptoms* SMQ (20000051), narrow and broad terms
   - For supraventricular arrhythmias: in *Cardiac arrhythmia* SMQ, under 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

TE ISRs will be summarized by treatment using the MedDRA PT in any of the following:

- HLT of *Injection site reaction*
- HLT of *Administration site reaction*
- 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 summarized by treatment groups by PT with decreasing frequency under following SMQs:

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

Hepatic Treatment-Emergent Adverse Events

TE potentially drug-related hepatic disorders will be summarized by treatment using the 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).

Metabolic Acidosis

Metabolic acidosis will be identified using the following MedDRA PTs:

Preferred Term	Code
Blood ketone body	10057593
Blood ketone body increased	10057594
Blood ketone body present	10057598
Diabetic ketoacidosis	10012671
Diabetic ketoacidotic hyperglycaemic coma	10012672
Diabetic ketosis	10012673
Euglycaemic diabetic ketoacidosis	10080061
Ketoacidosis	10023379
Ketonuria	10023388
Ketosis	10023391
Lactic acidosis	10023676
Urine ketone body	10059222
Urine ketone body present	10057597

Hypotension, Orthostatic Hypotension, and Syncope

Hypotension, orthostatic hypotension, and syncope will identified using the following MedDRA PTs:

Preferred Term	Code
Diastolic hypotension	10066077
Hypotension	10021097
Hypotensive crisis	10083659
Orthostatic hypotension	10031127
Blood pressure ambulatory decreased	10005731
Blood pressure decreased	10005734
Blood pressure diastolic decreased	10005737
Blood pressure orthostatic decreased	10053356
Blood pressure systolic decreased	10005758
Mean arterial pressure decreased	10026983
Blood pressure orthostatic	10053352
Dizziness	10013573
Presyncope	10036653
Syncope	10042772
Drop attacks	10013643
Loss of consciousness	10024855

5.3. Appendix 3: Cardiac Magnetic Resonance Imaging Substudy

This section is applicable to the participants who are enrolled in the cardiac magnetic resonance imaging (MRI) addendum.

This addendum applies to a subset of participants (approximately 150 participants) enrolled at selected sites that have the technical capability of conducting a cardiac MRI.

The participant demographics and baseline characteristics for the MRI substudy will be summarized for all participants enrolled in the MRI substudy.

Summaries of study disposition and study drug disposition will be provided for all participants enrolled in the MRI substudy.

MRI analyses will be guided by the treatment policy strategy and conducted for participants who are enrolled in the addendum, received at least 1 dose of study drug, and have MRI measurements at both baseline and postbaseline. The baseline MRI is defined as the MRI measurement taken prior or within 7 days of the second dose of the study treatment. The measurement for the Week 52 MRI will include a measurement taken either at 52 weeks, or within 105 days after 52 weeks, or at the early discontinuation visit if the discontinuation occurs prior to 52 weeks.

Objectives	Endpoints	Analytical Approaches
Evaluation of cardiac function and structure by cardiac MRI	Changes from baseline to Week 52 for the following: Structural and functional parameters • Left ventricular mass and index (LVM and LVMI, respectively) • Left ventricular end diastolic volume and index (LVEDV and LVEDVI, respectively) • Left ventricular end systolic volume and index (LVESV and LVESVI, respectively) • Left atrial volume and index (LAV and LAVI, respectively) • Left ventricular ejection fraction (LVEF) • Left ventricular cardiac output (LVCO) • Left ventricular stroke volume (LVSV) Feature tracking • Left ventricular global longitudinal strain (LVGLS) • Left ventricular global circumferential strain (LVGCS) • Left apical endocardial global longitudinal strain (LAEGLS) • Left apical endocardial global circumferential strain (LAEGCS) Adipose tissue volumes • Epicardial fat volume • Pericardial fat volume	Change from baseline to Week 52 for each parameter will be compared between treatment arms using an ANCOVA approach. The model will include treatment, the stratification factors of diagnosed T2DM (Y/N) and BMI group (<35, ≥35 kg/m^2), and the baseline value for the parameter. Summary statistics for MRI parameters at baseline and at Week 52 will be provided. The treatment comparison at baseline will be performed using an ANOVA model.

Abbreviations: ANCOVA = analysis of covariance; ANOVA = analysis of variance; BMI= body mass index; MRI = magnetic resonance imaging; T2DM = type 2 diabetes mellitus.

5.4. Appendix 4: Statistical Analysis for China

Analyses will be performed for the following subpopulation:

- participants enrolled in China (mainland China and Taiwan).

The analysis methods for this subpopulation will be similar to those described for the main part SAP GPID. If there is not a sufficient number of participants in the subpopulation, summary statistics will be provided.

The analyses to be included will be documented in a separate list of analyses which should include disposition, demographics, and selected efficacy and safety endpoints.

6. References

[ADA] American Diabetes Association. 6. Glycemic targets: Standards of medical care in diabetes - 2021. *Diabetes Care.* 2021:44(suppl 1):S73-S84. https://doi.org/10.2337/dc21-S006

Dong G, Qiu J, Wang D. The stratified win ratio. *Journal of Biopharmaceutical Statistics*, 2018; 778-796. https://doi.org/10.1080/10543406.2017.1397007

Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics.* 2000; 56(2):554-562. https://doi.org/10.1111/j.0006-341x.2000.00554.x

Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol*. 2000;35(5):1245-1255. https://doi.org/10.1016/s0735-1097(00)00531-3

Joseph SM, Novak E, Arnold SV, et al. Comparable performance of the Kansas City Cardiomyopathy Questionnaire in patients with heart failure with preserved and reduced ejection fraction. *Circ Heart Fail.* 2013;6(6):1139-1146. https://doi.org/10.1161/CIRCHEARTFAILURE.113.000359

Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B.* 2000;62:711–730. https://doi.org/10.1111/1467-9868.00259

Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J.* 2012;33(2):176-182. https://doi.org/10.1093/eurheartj/ehr352

PHUSE. Analysis and displays associated with adverse events: focus on adverse events in phase 2-4 clinical trials and integrated summary document. February 2017. Accessed June 24, 2024. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+and+Displays+Associated+with+Adverse+Events+Focus+on+Adverse+Events+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons Inc.; 1987.

Van Elteren, PH. On the combination of independent two sample tests of Wilcoxon. *Bull Inst Int Stats.* 1960;37 :351–361.

Reddy YNV, Carter RE, Sundaram V, Kaye DM, et al. An evidence-based screening tool for heart failure with preserved ejection fraction: The HFpEF-ABA score. *Nature Medicine*, 2024; in press

Signature Page for VV-CLIN-064990 v2.0

Approval	PPD 26-Jun-2024 12:10:19 GMT+0000

Signature Page for VV-CLIN-064990 v2.0