Statistical Analysis Plan I8F-MC-GPGM (V2)

# Efficacy and Safety of LY3298176 Once Weekly versus Insulin Glargine in Patients with Type 2 Diabetes and Increased Cardiovascular Risk (SURPASS-4)

NCT03730662

Approval Date: 15-Feb-2021

# 1. Statistical Analysis Plan I8F-MC-GPGM: Efficacy and Safety of Tirzepatide Once Weekly versus Insulin Glargine in Patients with Type 2 Diabetes and Increased Cardiovascular Risk (SURPASS-4)



LY3298176 for Type 2 Diabetes Mellitus

Phase-3 randomized 4-arm parallel design open label trial comparing 3-doses of LY3298176 to Insulin Glargine in patients with Type 2 Diabetes and increased cardiovascular risk


Eli Lilly and Company
Indianapolis, Indiana USA 46285


Protocol I8F-MC-GPGM
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 11-Dec-2018.
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly on date provided below.

Approval Date: 15-Feb-2021 GMT

# 2. Table of Contents

**Section** **Page**

## Table of Contents

**Table** **Page**

## Table of Contents

**Figure** **Page**

## Revision History

SAP Version 1 was approved prior to the first patient receiving study drug.

SAP version 2 was approved prior to the interim cardiovascular (CV) meta-analysis and final database lock. The main changes of this version are:

1. Updated definition of analysis set: Per agreement with the US Food and Drug Administration (FDA), patients that discontinue study drug due to inadvertent enrollment are excluded from efficacy analyses.
2. Updated baseline definition for selected measures based on scientific consideration and to minimize missing values at baseline.
3. Updated primary endpoint: Missing data imputation. Corrected the definition of "retrieved dropout." Use of local laboratory data for glycemic control measures when central laboratory data are not available.
4. Updated the language to handle lack of convergence in longitudinal logistic regression analysis due to low number of events for hemoglobin A1c (HbA1c) and weight loss target analyses.
5. Updated the language in section on adverse events of special interest and renamed section to "Special Safety Topics" to be consistent with the program SAP (see Section 5.13.2).
6. Updated the language in Section 5.13.2.1 to provide clarity on definitions of different categories of hypoglycemic events and analysis of hypoglycemic events.
7. Added Section 5.13.3 for additional analyses on renal safety based on a revised definition of renal impairment (eGFR<60 mL/min/1.73m2 or UACR >300 mg/g at baseline).
8. Added language in Section 5.19.1 according to protocol amendment (b) for an interim data extraction to support the program wide interim cardiovascular safety meta-analysis.
9. Added Section 5.20 for COVID-19 impact assessment.

# 3. Study Objectives

## 3.1. Primary Objectives

Primary objectives of the study are to demonstrate that QW tirzepatide 10 mg and/or 15 mg is noninferior to insulin glargine for change from baseline in HbA1c at 52 weeks.

## 3.2. Key Secondary Objectives Subject to Strong Type 1 Error Rate Control

Together with the primary objectives, the following secondary objectives are subjected to strong control of type 1 error rate (see Section 5.12.3).

- To demonstrate that QW tirzepatide 5 mg is noninferior to insulin glargine for change from baseline in HbA1c at 52 weeks.
- To demonstrate that QW tirzepatide 5 mg, 10 mg, and/or 15 mg is superior to insulin glargine for change from baseline in weight at 52 weeks.
- To demonstrate that QW tirzepatide 5 mg, 10 mg, and/or 15 mg is superior to insulin glargine for change from baseline in HbA1c at 52 weeks.
- To demonstrate that QW tirzepatide 5 mg, 10 mg, and/or 15 mg is superior to insulin glargine for the proportion of patients with HbA1c target values of <7.0% (53 mmol/mol) at 52 weeks.

## 3.3. Other Secondary and Exploratory Efficacy Objectives Not subject to Type 1 Error Rate Control

The following objectives are considered exploratory and hence not subjected to strong control of type 1 error rate.

To demonstrate that QW tirzepatide 5 mg, 10 mg, and/or 15 mg is superior to insulin glargine at 52 weeks relative to the following:

- the mean change in fasting serum glucose from baseline.
- the proportion of patients achieving an HbA1c target value of ≤6.5% (48 mmol/mol), and <5.7% (39 mmol/mol).
- mean change in 7-point self-monitored blood glucose (SMBG) profiles from baseline.
- proportion of patients who achieved weight loss of ≥5%, ≥10%, and ≥15% from baseline.

## 3.4. Safety Objectives

To compare the safety of tirzepatide 5 mg, 10 mg, and 15 mg to insulin glargine at 52 weeks, and at the end of the safety follow-up period, with respect to the following outcomes and measures:

- Treatment-emergent adverse events (TEAEs)
- Early discontinuations of study drug due to adverse events (AEs)
- Adjudicated deaths and nonfatal major cardiovascular (CV) events
- Adjudicated pancreatic AEs

- Medullary thyroid carcinoma (MTC), C-cell hyperplasia, and serum calcitonin
- Incidence of treatment-emergent (TE) tirzepatide anti-drug antibodies (ADA) and systemic hypersensitivity reactions
- Mean change in systolic and diastolic blood pressure and heart rate from baseline
- Occurrence of hypoglycemic events
- Incidence of initiation of rescue therapy for severe, persistent hyperglycemia

## 3.5. Pharmacokinetics

To characterize the pharmacokinetics (PK) of tirzepatide 5 mg, 10 mg, and 15 mg and evaluate the relationships between tirzepatide exposure and safety, tolerability, and efficacy measures.

# 4. Study Design

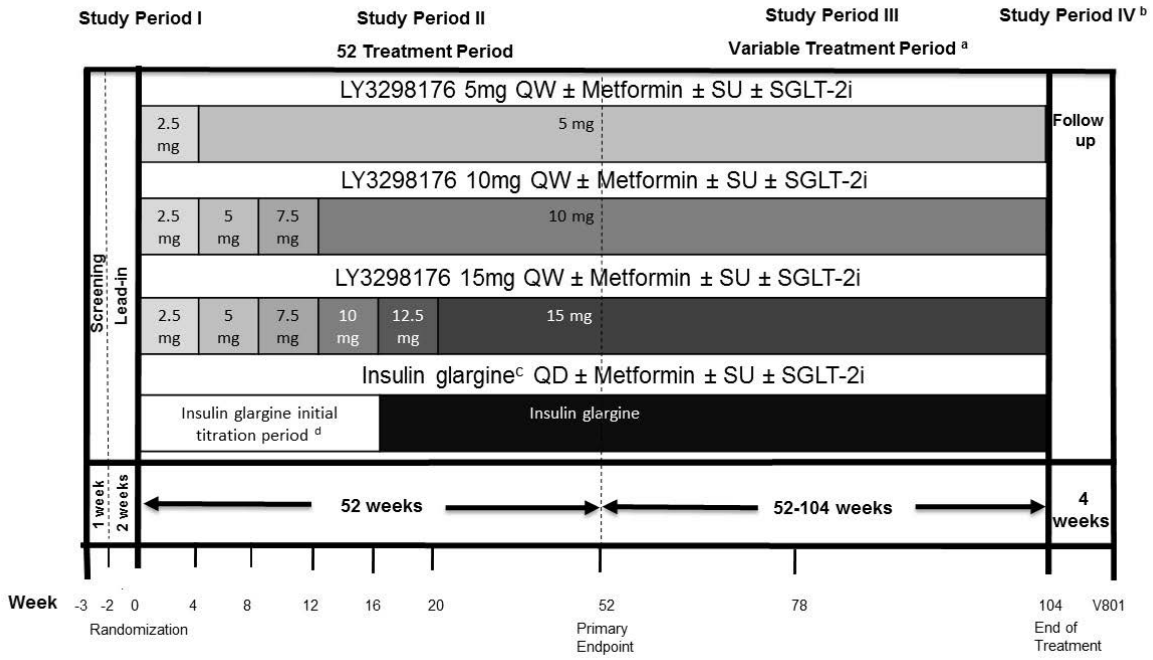
## 4.1. Summary of Study Design

Study GPGM is a Phase 3, open-label comparator, multicenter, parallel-arm, randomized study to compare the safety and efficacy of 3 maintenance doses of tirzepatide with titrated insulin glargine in patients with Type 2 Diabetes Mellitus with increased CV risk. The primary endpoint will be the mean change HbA1c from baseline to 52 weeks.

The study will continue until all of the following criteria are fulfilled:

1. At least 52 weeks from the time of the last patient randomized.
2. At least 300 patients assigned to the combined tirzepatide arms reach at least 78 weeks of treatment.
3. Approximately 110 patients in this study experience at least 1 component event of the composite CV endpoint of cardiovascular death, myocardial infarction, stroke, or hospitalization for unstable angina.

Figure GPGM.4.1 illustrates the study design.




Abbreviations: QD = once daily; QW = once weekly; SGLT-2i = sodium-glucose co-transporter-2 inhibitor; SU = sulfonylurea.

[a] Patients will be on study drug for at least 12 months and will receive no more than 24 months of treatment.

[b] All patients will perform a Visit 801 4 weeks after their last treatment visit.

[c] The starting dose of insulin glargine will be 10 IU/day at bedtime, titrated to a FBG <100 mg/dL, following a TTT algorithm (Riddle et al. 2003).

[d] Patients will titrate insulin glargine dose in a weekly manner and will make the dose decision with the investigator for the first 8 weeks (phone or clinic visit). From Week 8 to Week 16, patients will continue the titration by a phone consultation or clinic visit every other week, with 3 weeks between Visits 13 and 14.

**Figure GPGM.4.1. Illustration of study design for Clinical Protocol I8F-MC-GPGM.**

### Study Period I (screening and lead-in)

Screening (Visit 1)

The purpose of screening procedures at Visit 1 is to establish initial eligibility and to obtain blood samples for laboratory assessments needed to confirm eligibility at Visit 2. Patients who meet all applicable inclusion criteria and none of the applicable exclusion criteria at Visit 1 will continue on their prestudy therapy until Visit 2.

Lead-in (Visit 2 to Visit 3)

At Visit 2, the screening laboratory results will be reviewed and patient eligibility will be established with the exception of retinopathy status. Dilated fundoscopic exam will be performed between Visit 2 and Visit 3 as results are required to confirm eligibility.

**Study Period II (52-week treatment period)**

Randomization (Visit 3)

At Visit 3, patients will perform all required baseline study procedures (including the collection of all baseline laboratory measures) prior to randomization and prior to taking the first dose of study drug.

Postrandomization period (end of Visit 3 to Visit 24):

The starting dose of tirzepatide will be 2.5 mg QW for 4 weeks, followed by an increase to 5 mg QW, for the duration of the study low-dose arm. For the 10-mg arm, the starting dose of tirzepatide will be 2.5 mg QW for 4 weeks, then the dose will be increased by 2.5 mg every 4 weeks (2.5 to 5 to 7.5 to 10 mg) until the 10-mg dose is reached and maintained for the duration of the study. For the 15-mg arm, the starting dose of tirzepatide will be 2.5 mg QW for 4 weeks, then the dose will be increased by 2.5 mg every 4 weeks (2.5 to 5 to 7.5 to 10 to 12.5-15 mg) until the 15-mg dose is reached and maintained for the duration of the study.

The initial dose of insulin glargine will be 10 IU/day, ideally at bedtime, titrated to a FBG <100 mg/dL, following a treat-to-target (TTT) algorithm.

**Study Period III (variable treatment period)**

Long-term safety period (Visit 25 to Visit 29):

Some patients will continue to receive tirzepatide or insulin glargine for up to, but not longer than, 24 months, as determined by the sponsor.

When the sponsor determines that the study completion criteria have been met, all patients will return to the site for a final treatment visit (FTV) within approximately 30 days. Patients who attend Visit 29 (24 months) or the FTV are considered to have completed the treatment period.

**Study Period IV (safety follow-up period)**

Safety follow-up (Visit 801) visits:

All patients who complete the treatment period are required to complete Visit 801, a safety follow-up visit approximately 4 weeks after their last visit. Patients discontinuing the study early and performing an early termination (ET) visit will also be asked to perform the safety follow-up visit, so that the safety follow-up visit will be their final visit. During the safety follow-up period, patients will not receive study drug. Patients will be treated with another glucose lowering intervention decided upon by the investigator. Initiation of new antihyperglycemic therapy for the safety follow-up period will not be classified as “rescue therapy.”

## 4.2. Method of Assignment to Treatment

Approximately 1872 Patients who meet all criteria for enrollment will be randomized to one of the study treatment arms at Visit 3. Assignment to treatment arms will be determined by a computer-generated random sequence using an Interactive Web Response System. Patients will be randomized in a 1:1:1:3 ratio to receive 5 mg tirzepatide, 10 mg tirzepatide, 15 mg tirzepatide, or insulin glargine titrated to achieve a FPG <100mg/dL. The randomization will be stratified by country, baseline HbA1c concentration (≤8.5%, >8.5% [≤69, >69 mmol/mol]), and baseline SGLT-2i use (Yes or No).

# 5. A Priori Statistical Methods

## 5.1. Populations for Analyses

For purposes of analysis, Table GPGM.5.1 defines the following analysis sets:

**Table GPGM.5.1. Analysis Populations/Data Sets**

| Population/Data Set | Description |
|---|---|
| Screened population | All participants who sign informed consent |
| Randomized population | All patients who are randomly assigned a treatment arm |
| Modified intent-to-treat (mITT) population | All randomly assigned participants who are exposed to at least 1 dose of study drug. |
| Efficacy analysis set (EAS) | Data obtained during Study Period II and III from mITT, excluding patients discontinuing study drug due to inadvertent enrollment and data after initiating rescue antihyperglycemic medication or discontinuation of study drug (last dose date + 7 days). |
| Full analysis set (FAS) | Data obtained during Study Period II and III from mITT, excluding patients discontinuing study drug due to inadvertent enrollment, regardless of adherence to study drug or initiation of rescue antihyperglycemic medication. |
| Safety analysis set (SS) | Data obtained during Study Period II, III, and IV from mITT, regardless of adherence to study drug or initiation of new antihyperglycemic medication. |

## 5.2. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly) or its designee. All statistical analyses will be conducted with SAS Version 9.4 or higher unless otherwise stated. Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other changes to the data analysis methods described in the protocol, and the justification for making the change, will be described in the statistical analysis plan (SAP) or the clinical study report (CSR). Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (e.g., few events to justify conducting an analysis). Listing of events will be provided in such situations. Additional analyses of the data may be conducted as deemed appropriate without further changes made to the protocol or SAP, even after the primary or final database locks (DBL).

Patients inadvertently receiving incorrect study treatment are expected to switch to their randomized treatment arm as soon as possible. Patients assigned to tirzepatide may not be able to tolerate the maximum dose of the randomized treatment arm and may continue study participation on a reduced maintenance dose. Continuing on a reduced maintenance dose will neither be considered as discontinuation of randomized treatment nor will be considered as non-compliant to randomized treatment. Additionally, to avoid potential selection biases, unless stated otherwise, statistical summaries and analyses will be conducted based on randomized

maintenance dose regardless of the actual treatment received by the patient. Therefore participants will be analyzed according to the treatment they were randomized. In the event of a treatment error, participants will be analyzed according to the treatment they were randomized.

Unless specified otherwise, the last measurement during visit 1 to visit 3 (including unscheduled visits) collected prior to or on the first dose day will serve as baseline. For immunogenicity, data collected up to the first dose time will serve as baseline. For labs and ECG, baseline needs to be prior to or within one hour after the first dose time. For patient-reported outcome measures, data obtained at visit 3, regardless of the timing relative to first dose, will serve as baseline.

There will be 2 estimands of interest in evaluating primary and secondary efficacy objectives. First estimand, the "efficacy" estimand, represents efficacy prior to discontinuation of study drug without confounding effects of rescue therapy for persistent severe hyperglycemia. Analysis relative to" efficacy" estimand will be conducted using EAS. The second estimand, the "treatment-regimen" estimand, represents the efficacy irrespective of adherence to investigational product or introduction of rescue therapy for persistent severe hyperglycemia. Analysis relative to "treatment-regimen" estimand will be conducted using FAS.

Unless specified otherwise, safety analyses will be conducted relative to "treatment-regimen" estimand using the safety analysis set.

End of study participation for a patient will be the earliest of date of death, date of withdrawal from further participation in the study, or date of safety follow-up visit (visit 801). For patients considered to be lost-to-follow-up, end of study participation will be the date of lost-to-follow-up reported by the investigator. Patient data included in database after the last date of study participation (date of death, date of early termination, or date of safety follow-up) will be excluded from statistical analysis. Listing of such data may be provided.

Summary statistics for categorical measures (including categorized continuous measures) will include sample size, frequency, and percentages. Summary statistics for continuous measures will include sample size, mean, standard deviation (SD), median, minimum, and maximum. The summary statistics will be presented by the nominal visit.

Statistical treatment comparisons will only be performed between tirzepatide doses and insulin glargine. Since the trials are not adequately powered to detect differences among tirzepatide doses, comparisons among tirzepatide arms will not be performed unless otherwise specified.

Statistical summaries and results of statistical analysis will be displayed in the following treatment order: 5mg tirzepatide, 10mg tirzepatide, 15mg tirzepatide, and insulin glargine.

## 5.3. Adjustments for Covariates

The study is stratified by country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), and baseline SGLT2 inhibitor use (Yes or No). Where necessary to be included as a stratification factor, countries with fewer than 10 randomized patients will be pooled into one category (pooled country). For HbA1c related analyses, country/pooled country and baseline SGLT2 inhibitor use will be used as stratification factors and baseline HbA1c as a covariate. For other efficacy analyses, country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), and baseline SGLT2 inhibitor use (Yes or No) will be used as stratification factors and respective baseline value as a covariate.

Stratification factors will be derived based on the data collected from the clinical database.

## 5.4. Handling of Dropouts or Missing Data

For the primary and key secondary efficacy endpoint analyses subject to type 1 error rate control, data for patients with missing values at the 52-week visit will be imputed based on the method describe in Section 5.12.1.3. Unless specified otherwise, imputation of missing data will be limited to primary and key secondary efficacy endpoint analysis. Missing other secondary or exploratory efficacy parameter values and missing safety laboratory values will not be explicitly imputed.

## 5.5. Multicenter Studies

To investigate potential regional influence on efficacy, country/pooled country will be used as a stratification factor in primary and secondary efficacy analysis.

## 5.6. Multiple Comparisons/Multiplicity Adjustments

Type 1 Error rate control strategy for primary and key secondary efficacy objectives is illustrated in Section 5.12.3. As they are intended for different purposes, no multiplicity adjustments will be made for conducting separate analyses relative to "efficacy" and "treatment-regimen" estimands. No multiplicity adjustments will be made for evaluating other secondary and exploratory efficacy objectives and safety assessments.

## 5.7. Patient Disposition

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

A listing of final study disposition and a listing of randomized treatment assignment (planned treatment) for all randomized patients will be provided. Final study disposition and treatment disposition for all randomized patients will be summarized by planned study treatment. Patient study disposition prior to the 52-week visit (when the primary endpoint is ascertained) as well as at or after the 52-week visit will be summarized by planned study treatment.

## 5.8. Patient Characteristics

Listing of patient demographics will be provided. All demographic and baseline clinical characteristics will be summarized by study treatment for the patients in the mITT population. Baseline demographic and clinical characteristics of special interest include: age, gender, race, ethnicity, weight, country of enrollment, HbA1c, fasting serum glucose, duration of type 2 diabetes, baseline antihyperglycemic medication, CV history, estimated glomerular filtration rate (eGFR), urine albumin-to-creatinine ratio (UACR), results of the fundoscopic exam, history of gallbladder disease, and percentage of patients who have renal impairment (defined as baseline eGFR<60 mL/min/1.73$m^2$ or baseline UACR>300 mg/g).

## 5.9. Concomitant Therapy

The prespecified concomitant medications of interest will be summarized by treatment at randomization. Additionally medications of interest initiated after randomization and change to medications of interest used at randomization will be summarized. The concomitant therapies will be mapped using the World Health Organization (WHO) DRUG dictionary in the clinical trial database and will be further classified using Anatomic-Therapeutic-Chemical (ATC) codes for reporting purposes.

The concomitant medications of interest include the following groups of medication:

- Baseline Antihyperglycemic therapy
    - Prior GLP-1 RA use by type
    - Metformin
    - Sulfonylureas, by type
    - SGLT2-inhibitors, by type
- Baseline antihypertensive therapy, by type
- Baseline lipid lowering therapy, by type
- Changes to baseline medication
    - Antihyperglycemic therapy
    - Antihypertensive therapy
    - Lipid lowering therapy
- Rescue therapy
- Initiation of following medications in Study Period II/III:
    - Antidiarrheal medication
    - Antiemetic medication

## 5.10. Treatment Exposure and Compliance

Listing of patients randomized but not receiving study treatment will be provided, if applicable. The listing will include patient identification, randomized treatment arm, and the reason for not receiving study treatment.

Summary of duration of follow-up (defined as time in days from date of randomization to date of safety follow-up) and 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 therapy.

### *5.10.1. Exposure and Compliance to Tirzepatide*

Number of patients prematurely discontinuing study treatment prior to the 52-week visit as well as before the end of efficacy follow-up will be provided by study treatment. Reasons for prematurely discontinuing study treatment prior to the 52-week visit as well as before the end of efficacy follow-up will be provided by study treatment. Time-to-event analysis of premature study treatment discontinuation will be conducted.

Proportion of patients with missing dosing information, receiving no tirzepatide dose, receiving 2.5mg, 5mg, 7.5mg, 10mg, 12.5mg or 15mg will be presented by randomized treatment and week from first dose.

Listing and summary of patients continuing on a reduced maintenance dose of tirzepatide compared to the randomized dose may be provided.

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

### *5.10.2. Exposure and Compliance to Insulin Glargine Treat-to-Target Algorithm*

Summary information of total daily dose of insulin glargine will be reported by visit. Information related to compliance to treat-to-target therapy including reasons for non-compliance will be summarized.

Compliance will be defined as taking at least 75% of the scheduled insulin doses. Compliance will be calculated by taking the number of injections expected minus the total number of injections missed (regardless of the actual dose administered) divided by the total number of injections expected to be administered × 100. Treatment compliance will be summarized descriptively using the mITT population.

## 5.11. Important Protocol Deviations

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

## 5.12. Efficacy Analysis

For the FDA and potentially for other regulatory agencies, all efficacy assessments will be guided by the "treatment-regimen" estimand conducted using FAS. Assessment of the primary and secondary efficacy objectives subject to type 1 error rate control (key secondary) will be conducted with multiple imputation of missing data (see Section 5.12.1.3) at 52 weeks. Assessment of other efficacy objectives will be conducted without imputation of missing data. For publications and other purposes, the assessment of efficacy objectives will be guided by the "efficacy" estimand using EAS set without imputation of missing data. A listing of patients randomized but not included in efficacy analyses (ie, not treated, discontinued treatment due to inadvertent enrollment) will be provided.

### *5.12.1. Primary Efficacy Analysis*

The primary efficacy measure will be change in HbA1c from baseline (post-baseline – baseline) at 52 weeks. Both HbA1c values as well as change from baseline in HbA1c will be summarized by treatment and nominal visit (week). When applicable, HbA1c data from a local lab will be used when central lab data is not available. If scheduled HbA1c data at the primary endpoint visit is not available, unscheduled HbA1c data collected for the primary endpoint visit will be included in analysis.

#### 5.12.1.1. The Analysis Relative to the Efficacy Estimand

The analysis will be conducted utilizing HbA1c data in EAS from baseline through the 52-week visit with the aid of a mixed model for repeated measures (MMRM). Restricted maximum likelihood (REML) will be used to obtain model parameter estimates and Kenward-Roger option will be used to estimate denominator degrees of freedom. The response variable of the MMRM model will be the primary measure and model terms will include treatment, visit, treatment by visit interaction, country/pooled country, and SGLT2 inhibitor use at baseline as fixed effects, and baseline HbA1c as a covariate. An unstructured covariance structure will be used to model the within-patient errors. If this model fails to converge, the following covariance structures will be tested in order:

- Heterogeneous Toeplitz;
- Heterogeneous First Order Autoregressive;
- Heterogeneous Compound Symmetry;
- Toeplitz;
- First Order Autoregressive;
- Compound Symmetry.

The first covariance structure that converges will be used. Resulting Least Squares Mean (LSM) estimate of mean change from baseline in HbA1c will be plotted by visit and by study treatment.

With the aid of the MMRM analysis, 2-sided 97.5% Confidence Interval (CI) for mean change in HbA1c from baseline to 52-week visit for (10mg tirzepatide – insulin glargine) as well as for

(15mg tirzepatide – insulin glargine) will be derived. If the upper limit of the CI is <= 0.3%, then the respective dose of tirzepatide (10mg and/or 15mg) will be declared noninferior to insulin glargine relative to change in HbA1c from baseline.

#### 5.12.1.2. The Analysis Relative to the Treatment-Regimen Estimand

The analysis will be conducted utilizing HbA1c data in FAS at baseline and at the 52-week visit with the aid of an Analysis of Covariance (ANCOVA). The response variable will be the primary measure and model terms will include treatment, country/pooled country, SGLT2 inhibitor use at baseline as fixed effects, and baseline HbA1c as a covariate. The ANCOVA analysis will be conducted with multiple imputation of missing primary measure (see Section 5.12.1.3 for details) and statistical inference over multiple imputation of missing data guided by Rubin (1987).

With the aid of the ANCOVA analysis, 2-sided 97.5% CI for mean change in HbA1c from baseline to 52-week visit for (10mg tirzepatide – insulin glargine) as well as for (15mg tirzepatide – insulin glargine) will be derived. If the upper limits of the CI are <= 0.3%, then the respective dose of tirzepatide (10mg and/or 15mg) will be declared noninferior to insulin glargine relative to change in HbA1c from baseline.

#### 5.12.1.3. Methods for Multiple Imputations

For efficacy analysis relative to "treatment-regimen" estimand, missing HbA1c data at the 52-week visit will be imputed based on "retrieved dropouts," defined as patients who had their HbA1c value measured at the 52-week visit in the same treatment arm who prematurely discontinued study drug. If the imputed value of HbA1c change from baseline is <-6.0% or >6.0%, that value will be set to -6.0% or 6.0%, respectively, to avoid unrealistic imputed values.

#### 5.12.1.4. Additional Analyses of the Primary Outcome

Upon successfully establishing noninferiority of tirzepatide compared to insulin glargine, superiority of tirzepatide compared to insulin glargine relative to change in HbA1c from baseline will be evaluated (see Section 5.12.3).

### *5.12.2. Secondary Efficacy Analyses Subject to Type 1 Error Rate Control*

#### 5.12.2.1. Mean change in HbA1c from baseline at the 52 week visit

Noninferiority of 5mg tirzepatide to insulin glargine will be conducted in a manner similar to Section 5.12.1. Assessment of Superiority of tirzepatide doses compared to insulin glargine will be conducted using the same statistical models as those used for evaluating the primary objective in Section 5.12.1. Decisions will be guided by the two-sided p-values for mean comparisons between tirzepatide doses and insulin glargine (see details in Section 5.12.3).

#### 5.12.2.2. Mean change in body weight from baseline at the 52 week visit

The analysis for change in body weight from baseline (postbaseline - baseline) will be conducted in a manner similar to the primary analysis in Section 5.12.1. Baseline HbA1c concentration (≤8.5%, >8.5% [≤69, >69 mmol/mol]) will be used as a fixed factor in place of baseline HbA1c as a covariate and baseline body weight will be used as an additional covariate in the statistical

model. Least Squares Mean (LSM) estimate of mean change in body weight from baseline will be plotted by nominal visit and by study treatment. For multiple imputation of missing values, if the imputed value of weight change from baseline is <-50 kg or >50 kg, then that value will be set to -50 kg or 50 kg, respectively, to avoid unrealistic imputed values.

#### 5.12.2.3. Proportion of Patients Achieving HbA1c<7% at the 52 week Visit

The analysis relative to the "efficacy" estimand for the endpoint at 52 weeks will be conducted using EAS with missing values imputed from an MMRM model and then dichotomized. The MMRM model includes country/pooled country, SGLT2 inhibitor use at baseline, treatment, visit, and treatment-by-visit interaction as fixed effects, and baseline HbA1c as a covariate. After dichotomizing continuous HbA1c, the data is analyzed using a logistic regression model with treatment, country/pooled country, SGLT2 inhibitor use at baseline, and baseline HbA1c as a covariate. In addition, analysis will be conducted utilizing data using EAS from baseline through the 52-week visit with the aid of a longitudinal logistic regression with repeated measurements with country/pooled country, SGLT2 inhibitor use at baseline, treatment, visit, treatment-by-visit interaction as fixed effects, and baseline HbA1c as a covariate. In the case that longitudinal logistic model does not converge due to small number of events, logistic regression will be utilized to analyze proportion of patients achieving HbA1c<7% at nominal visits.

Analysis relative to "treatment-regimen" estimand will be conducted utilizing HbA1c data in FAS at baseline and at the 52-week visit with the aid of a logistic regression with multiple imputation of missing HbA1c data at the 52-week visit (see Section 5.12.1.3 for details). Model terms will include treatment, country/pooled country, SGLT2 inhibitor use at baseline as fixed effects, and baseline HbA1c as a covariate and statistical inference over multiple imputations will be guided by Rubin (1987).

### *5.12.3. Type 1 Error Rate Control Strategy for Primary and Key Secondary Efficacy Analysis.*

Since they are intended for different purposes, no type 1 error rate adjustments will be made for conducting analysis relative to "efficacy" and "treatment-regimen" estimands. For analysis within each estimand, type 1 error rate control strategy for evaluation of primary and key secondary objectives is illustrated in Figure GPGM.5.1.

As illustrated in the figure,

1. $H_{15,1}$, $H_{15,2,}$ and $H_{15,3}$ are evaluated hierarchically each at two-sided 0.025 significance level conditioned on the successfully achieving the preceding objective. In parallel,

2. $H_{10,1}$, $H_{10,2,}$ and $H_{10,3}$ are evaluated hierarchically each at two-sided 0.025 significance level conditioned on the successfully achieving the preceding objective.

3.

   a) If all objectives in #1 and #2 above are successfully established, $H_{5,1}$, $H_{5,2,}$ and $H_{5,3}$ are evaluated hierarchically, each at two-sided 0.05 significance level.

b) If all objectives in only #1 or only #2 above are successfully established, $H_{5,1}$, $H_{5,2,}$ and $H_{5,3}$ are evaluated hierarchically, each at two-sided 0.025 significance level.

4. If all objectives: $H_{5,1}$, $H_{5,2,}$ and $H_{5,3}$ are successfully established and

a) if all objectives in #1 and #2 above are successfully established, then $H_{10,4}$ , $H_{15,4}$ and $H_{5,4}$ will be evaluated hierarchically each at two-sided 0.05 significance level conditioned on the successfully achieving the preceding objective.

b) if all objectives in only #1 or only #2 above are successfully established, then $H_{10,4}$, $H_{15,4}$ and $H_{5,4}$ will be evaluated hierarchically each at two-sided 0.025 significance level conditioned on the successfully achieving the preceding objective.

5. If all objectives in #3 and #4 above are successfully established, and at least 1 objective from #1 or #2 above is not successfully established, recycle 100% of the unused alpha back to #1 or #2 above.

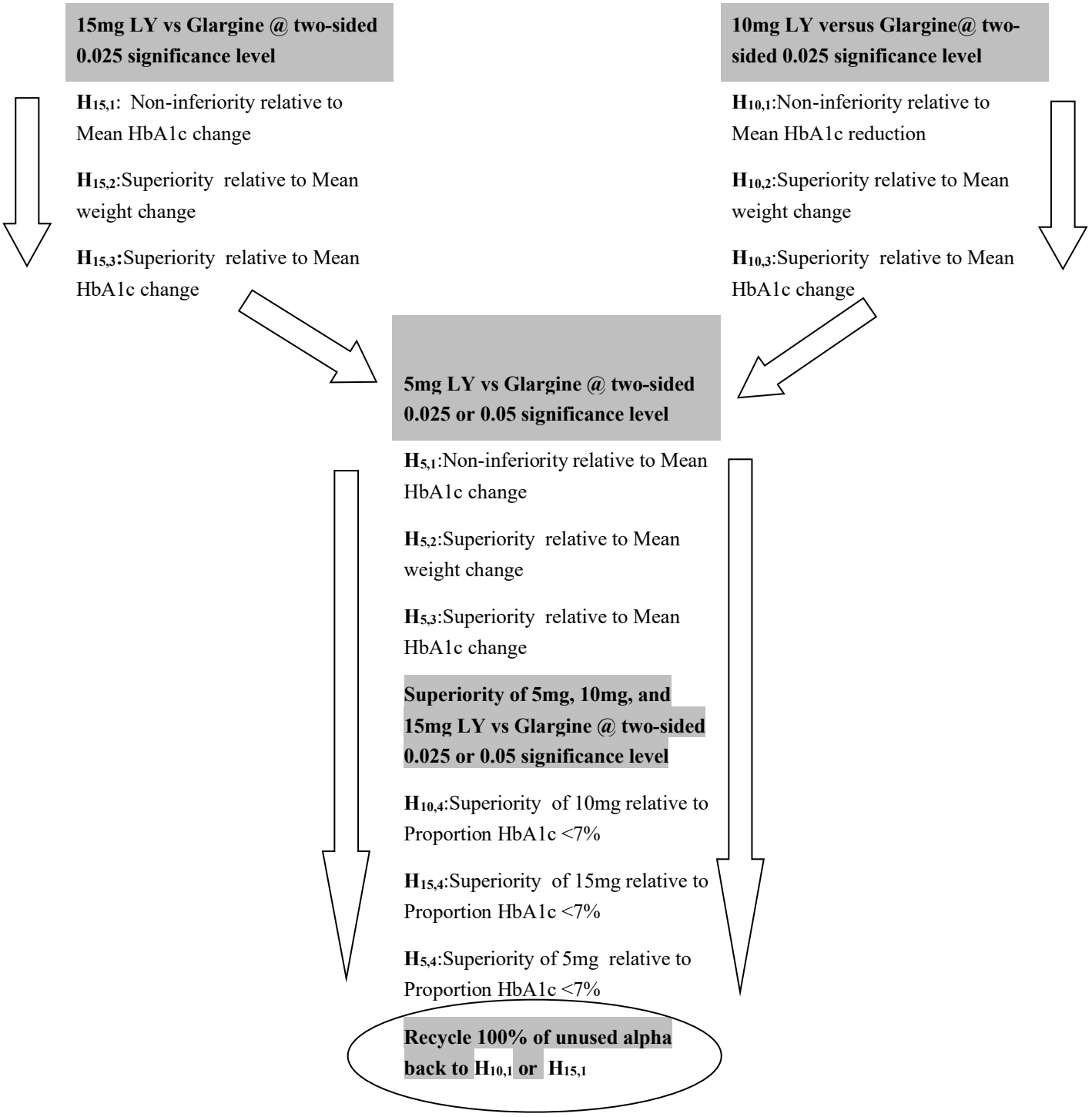



**Figure GPGM.5.1. Type 1 Error control strategy for primary and key secondary efficacy endpoints.**

### *5.12.4. Other Secondary and Exploratory efficacy Analyses*

Other secondary and Exploratory efficacy measures will be summarized by treatment and nominal visit. Statistical analyses will be conducted in a manner similar to Sections 5.12.1 and 5.12.2. However, missing data will not be imputed and assessments are not subject to type 1 error rate control.

| Objective | Relative to the efficacy measure: | Analysis Conducted in a manner similar to: | Additional Information |
|---|---|---|---|
| QW tirzepatide 5 mg, 10 mg, and/or 15 mg is superior to insulin glargine at 52 weeks | Change from baseline in fasting serum glucose | 5.12.1 | Use baseline HbA1c strata as a fixed effect in place of baseline HbA1c as a covariate. Use baseline FSG as a covariate. LSM estimates will be plotted by treatment and visit. |
| | Change from baseline in 7-point self-monitored blood glucose (SMBG) profiles | 5.12.1 | Use baseline HbA1c strata as a fixed effect in place of baseline HbA1c as a covariate. Use baseline SMBG as a covariate. LSM estimates at 52-weeks will be plotted by treatment and 7-points. |
| | proportion of patients achieving an HbA1c target value of ≤6.5% (48 mmol/mol), | 5.12.2.3 | None |
| | proportion of patients achieving an HbA1c target value of <5.7% (39 mmol/mol). | 5.12.2.3 | None |
| | proportion of patients who achieved weight loss of ≥5%, from baseline | 5.12.2.3 | Use baseline HbA1c strata as a fixed effect in place of baseline HbA1c as a covariate and baseline weight as a covariate |
| | proportion of patients who achieved weight loss of ≥10% from baseline | 5.12.2.3 | Use baseline HbA1c strata as a fixed effect in place of baseline HbA1c as a covariate and baseline weight as a covariate |
| | proportion of patients who achieved weight loss of ≥15% from baseline | 5.12.2.3 | Use baseline HbA1c strata as a fixed effect in place of baseline HbA1c as a covariate and baseline weight as a covariate |
| To compare tirzepatide 5 mg, 10 mg, and 15 mg to insulin glargine at 52 weeks | Change from baseline in waist circumference | 5.12.1 | Use baseline HbA1c strata as a fixed effect in place of baseline HbA1c as a covariate. Use baseline waist circumference as a covariate. |
| | Change from baseline in lipid parameters (Total-Cholesterol, HDL, VLDL, TG) | 5.12.1 | Use baseline HbA1c strata as a fixed effect in place of baseline HbA1c as a covariate. Use corresponding |

| | | | |
|---|---|---|---|
| | | | baseline lipid parameter as a covariate. |
| To characterize long-term (beyond 52 weeks) glycemic control with tirzepatide 5 mg, 10 mg, and 15 mg | Change from baseline in HbA1c | 5.12.1 | The analysis will be conducted utilizing HbA1c data from baseline through the end of efficacy follow-up |
| QW tirzepatide 5 mg, 10 mg, and/or 15 mg is superior to insulin glargine at 52 weeks | Proportion of patients achieving HbA1c target ≤6.5% and without weight gain (<0.1 kg) and without documented symptomatic hypoglycemia or severe hypoglycemia | 5.12.2.3 | Include rate of hypoglycemic events at baseline and baseline body weight as additional covariates |
| | Proportion of patients achieving HbA1c target < 7.0% and without weight gain (<0.1 kg) and without documented symptomatic hypoglycemia or severe hypoglycemia | 5.12.2.3 | Include rate of hypoglycemic events at baseline and baseline body weight as additional covariates |
| | Proportion of patients achieving HbA1c target ≤6.5% and without weight gain (<0.1 kg) and without clinically significant documented symptomatic hypoglycemia or severe hypoglycemia | 5.12.2.3 | Include rate of hypoglycemic events at baseline and baseline body weight as additional covariates |
| | Proportion of patients achieving HbA1c target < 7.0% and without weight gain (<0.1 kg) and without clinically significant documented symptomatic hypoglycemia or severe hypoglycemia | 5.12.2.3 | Include rate of hypoglycemic events at baseline and baseline body weight as additional covariates |

## 5.13. Safety Analysis

Unless specified otherwise, safety assessments will be based on the Safety analysis Set (see Table GPGM.5.1). All events that occur between the date of first dose of study drug to the date of patient's safety follow-up visit or patient's end of study participation will be included, regardless of the adherence to study drug or initiation of rescue therapy. For assessing benefit and risk profile through 52-weeks, selected safety analyses will be conducted by utilizing safety data from first dose through the date of 52 week visit. Some safety analysis may be conducted after excluding data after the initiation of new antihyperglycemic therapy. Unless specified otherwise, statistical assessment of homogeneity of the distribution of categorical safety responses among treatment arms will be conducted using Fisher's exact test.

Unless specified otherwise, difference among treatment mean change from baseline in continuous safety parameters at all scheduled visits will be assessed via a MMRM using REML. The model will include country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), baseline SGLT2 inhibitor use (Yes or No), treatment group, visit, and treatment-by-visit interaction as fixed effects, and baseline value of the safety parameter as a covariate. To model the covariance structure within patients, the unstructured covariance matrix will be used. If this model fails to converge, the following covariance structures will be tested in order specified in Section 5.12.1.1.

For selected safety parameters, time-to-first-event analysis via Cox proportional hazards model may be conducted. For patients without event, "time-to-event" will be censored as time (in days) from first dose to end of study participation (date of death, study discontinuation, or date of safety follow-up). For patients experiencing the event, "time-to-first-event" will be the time (in days) from first dose to first occurrence of the event.

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

### *5.13.1. Adverse Events*

A listing of AEs occurring either before first dose or after patient's last date of study participation will be provided. Listing will include patient identification including the treatment, site number, event information: AE group ID, event start date, MedDRA System Organ Class (SOC), and Preferred Term (PT), seriousness, severity, outcome, relationship to study drug, time from first dose of study drug to the event, time from last dose of study drug to event, and time from end of study participation to the event.

A TEAE is defined as an event that first occurred or worsened in severity after baseline. 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, it will be treated as 'mild' in severity for determining treatment-emergence. Events with a missing severity during the postbaseline period will be treated as 'severe' and treatment-emergence will be determined by comparing to baseline severity.

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

Overview of the number and percentage of patients who experienced a TEAE, serious adverse event (SAE), death, discontinued from study treatment, or study due to an AE, and relationship to study drug, will be summarized by treatment.

The percentages of patients with TEAEs, overall and common (common TEAEs occurred in >=5% of any treatment arm before rounding of treated patients), will be summarized by treatment using MedDRA PT. Events will be ordered by decreasing frequency.

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

Patient narratives will be provided for all patients who experience any of the following events:

- Deaths
- SAEs
- Permanent discontinuations of study treatment due to AEs
- Severe adverse events of special interest

#### 5.13.1.1. Deaths

A listing of all deaths will be provided. Listing will include patient identification including the treatment, site number, date of death, age at the time of enrollment, sex, MedDRA PT of associate AE, time from first dose of study drug to death, time from last dose of study drug to death (if patient had discontinued study drug), cause of death as reported by investigator, caused of death as adjudicated by clinical endpoint committee (CEC).

#### 5.13.1.2. Other Serious Adverse Events

The number and percentage of patients who experienced a SAE (including deaths and SAEs temporally associated or preceding deaths) during the study follow-up will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC.

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

#### 5.13.1.3. Discontinuation from Study Due to Adverse Event

The number and percentage of patients who prematurely discontinue study due to AE will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC.

#### 5.13.1.4. Discontinuation from Study Drug Due to Adverse Event

The number and percentage of patients who prematurely discontinue study drug due to AE will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency within SOC. Time-to-event analysis of time to study drug discontinuation by treatment as well as time to study drug discontinuation due to AE will be conducted.

## *5.13.2. Special Safety Topics*

### 5.13.2.1. Hypoglycemic Events

Definitions of different categories of hypoglycemic events are as follows:

| | **Symptoms and/or Signs of hypoglycemia** | **Blood Glucose Level** |
|---|---|---|
| **Glucose Alert Value:** | | |
| Documented symptomatic hypoglycemia | Yes | ≤70 mg/dL (3.9 mmol/L) |
| Documented asymptomatic hypoglycemia | No | ≤70 mg/dL (3.9 mmol/L) |
| Documented unspecified hypoglycemia | Unknown | ≤70 mg/dL (3.9 mmol/L) |
| **Clinically Significant Hypoglycemia:** | | |
| Clinically significant Documented symptomatic hypoglycemia | Yes | <54 mg/dL (3.0 mmol/L) |
| Clinically significant Documented asymptomatic hypoglycemia | No | <54 mg/dL (3.0 mmol/L) |
| Clinically significant Documented unspecified hypoglycemia | Unknown | <54 mg/dL (3.0 mmol/L) |
| **Severe Hypoglycemia** | | |

**Severe hypoglycemia**: Defined as an episode with severe cognitive impairment requiring the assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions. Severe hypoglycemia will be reported as a SAE. Severe hypoglycemia will be considered an adverse event of special interest (AESI).

**Nocturnal hypoglycemia**: Defined as any hypoglycemic event that occurs between bedtime and waking.

To avoid duplicate reporting, all consecutive hypoglycemic events in the same category, occurring within a 1-hour period may be considered to be a single hypoglycemic event.

Statistical summaries and analyses will exclude hypoglycemia events occurring after initiation of a new antihyperglycemic therapy. Severe hypoglycemia and hypoglycemia incidence (with blood glucose level ≤70 mg/dL [3.9 mmol/L] or <54 mg/dL [3.0 mmol/L], separately) as well as rate per patient year of exposure will be provided by treatment. The incidence of hypoglycemic event will be analyzed using logistic regression with treatment, country/pooled country, SGLT2i use at baseline (Yes/No) and baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]) as the fixed effects. The rate of hypoglycemic episodes per patient year will be analyzed using a generalized

linear mixed-effects model assuming the number of hypoglycemic episodes follow a negative binomial distribution with mean modeled using country/pooled country, SGLT2i use at baseline (Yes/No), baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]) and treatment as fixed effects. The logarithm of days during analysis interval will be adjusted as an offset to account for possible unequal treatment duration of follow up between patients.

Summary of hypoglycemic events will also be provided in subjects on or not on stable dose of sulfonylureas, separately.

#### 5.13.2.2. Severe Persistent Hyperglycemia

A summary or listing of initiation of rescue therapy in response to severe, persistent hyperglycemia will be provided by treatment. If there is sufficient number of episodes, time-to-first-event analysis for the initiation of rescue therapy will be conducted by treatment using Cox proportional regression model. A listing of patients initiating rescue therapy will be provided.

#### 5.13.2.3. Pancreatitis

If data warrants, summaries of adjudicated and investigator-reported pancreatic events will be provided by treatment. Determination of investigator-reported events will be through the predefined Standardized MedDRA Queries (SMQ) search for acute pancreatitis and MedDRA PT of pancreatitis chronic. Detailed searching criteria can be found in Appendix 1. Treatment-emergent adjudication-confirmed pancreatitis will be considered an AESI.

##### *5.13.2.3.1. Pancreatic Enzyme Assessment*

Observed pancreatic enzyme data (p-amylase and lipase) will be summarized by treatment and nominal visit. Additionally, number and proportion of patients with maximum postbaseline pancreatic enzyme value exceeding the following thresholds will be provided by maximum baseline pancreatic enzyme value (≤ 1×Upper Limit of Normal [ULN], > 1×ULN) and treatment: ≤ 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 (post-baseline measure/baseline measure) response variable, and stratification factors, treatment, nominal visit, and treatment-by-nominal visit interaction as fixed effects, and baseline value as a covariate.

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

Treatment-emergent thyroid malignancies and C-cell hyperplasia will be identified using a pre-defined MeDRA HLTs of thyroid neoplasms malignant and PT of thyroid C-cell hyperplasia. Detailed searching criteria can be found in Appendix 1. A summary by treatment and PT and a listing will be provided. Thyroid malignancies and C-cell hyperplasia will be considered as AESI.

#### 5.13.2.5. Malignancies

The AE database will be searched using pre-defined SMQs to identify events consistent with malignancy. Detailed searching criteria can be found in Appendix 1. A summary by treatment and PT and a listing will be provided. Malignancy will be considered as AESI.

#### 5.13.2.6. Calcitonin

Observed calcitonin data will be summarized by treatment and nominal visit. Additionally, number and proportion of patients with a maximum postbaseline calcitonin value exceeding the following thresholds will be provided by treatment, maximum baseline calcitonin value (≤20ng/L, >20 to ≤35ng/L, >35ng/L),: ≤20 ng/L, >20 to ≤35 ng/L, >35 to ≤50 ng/L, >50 to ≤100 ng/L, >100ng/L.

#### 5.13.2.7. Major Adverse Cardiovascular Events (MACE)

Major adverse cardiovascular events reported by investigators are adjudicated by an independent clinical endpoint committee (CEC) in a blinded fashion. The MACE events of special interest are: deaths due to cardiovascular cause, myocardial infarction, hospitalization for unstable angina, hospitalization for heart failure, coronary interventions (such as coronary artery bypass graft [CABG] or percutaneous coronary intervention [PCI]); and cerebrovascular events, including cerebrovascular accident (stroke) and transient ischemic attack (TIA). Deaths with an undetermined cause per CEC will be included in death due to CV cause for analysis purposes.

A listing of patients reporting MACE events, either reported by investigator or identified by the CEC, will be provided. Listing will include treatment, patient identification including the site number, date of event, type of event as reported by the investigator, type of event as adjudicated by the CEC, time from first dose of study drug to the event, time from last dose to the event (if patient has discontinued study drug prior to the event). Only positively adjudicated MACE will be considered as an AESI.

Summaries of positively CEC adjudicated and investigator-reported MACE events will be provided by treatment and event type.

Time-to-first-event analysis of the composite endpoint consisting of positively CEC adjudicated: death due to cardiovascular cause, myocardial infarction, stroke, or hospitalization due to unstable angina (MACE-4) will be conducted.

#### 5.13.2.8. Supraventricular Arrhythmias and Cardiac Conduction Disorders

Adverse event database will be searched using a pre-defined MeDRA terms to identify events consistent with supraventricular arrhythmias and cardiac conduction Disorders. Detailed searching criteria can be found in Appendix 1. Incidence of resulting TEAEs will be summarized by treatment and PT within SMQ and HLT. Treatment-emergent severe/serious supraventricular arrhythmias and cardiac conduction disorders will be considered AESIs.

#### 5.13.2.9. Hypersensitivity Events

Two main analyses are performed for hypersensitivity reactions and related information:

a. **Potential Immediate Hypersensitivity**: Analysis of TEAE occurring from start of study drug administration up to 24 hours after end of study drug administration. For events without the hypersensitivity eCRF, only date (no time) information is collected, the events occurred on the same date as the study drug injection date will be included.

b. **Potential Non-Immediate Hypersensitivity:** Analysis of TEAE occurring more than 24 hours after end of study drug administration, but prior to subsequent study drug administration.

Summaries of all potential hypersensitivity reactions will be generated by PT with decreasing frequency by treatment. The AE database will be searched using predefined SMQs to identify events consistent with hypersensitivity events. Detailed searching criteria for hypersensitivity events can be found in Appendix 1. Severe/serious hypersensitivity events identified by predefined SMQ searches will be considered AESIs.

#### 5.13.2.10. Injection Site Reactions

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

Additionally, potential injection site reactions will be searched by pre-defined MedDRA High Level Term (HLT) of injection site reactions and infusion related reactions. Detailed searching criteria for hypersensitivity events can be found in Appendix 1. The PT will be used for summary within each HLT category. Only severe/serious injection site reactions will be considered AESIs.

#### 5.13.2.11. Immunogenicity

Treatment-emergent anti-drug antibodies (TE ADA) are defined as those with a titer 2-fold (1 dilution) greater than the minimum required dilution if no ADAs were detected at baseline (treatment-induced ADA) or those with a 4-fold (2 dilutions) increase in titer compared to baseline if ADAs were detected at baseline (treatment boosted ADA). A patient is evaluable for TE ADA if the patient has a non-missing baseline ADA result, and at least 1 non-missing postbaseline ADA result.

The frequency and percentage of patients with preexisting ADA, with TE ADA, and with neutralizing TE ADA to tirzepatide will be tabulated by dose, where proportions are relative to the number of patients who are TE ADA evaluable.

A listing may be provided of all immunogenicity assessments for those patients who at any time had TE ADA Present. This includes the tirzepatide concentration from a simultaneous PK sample, and the clinical interpretation result .

Depending on the number of patients with TE ADA, selected efficacy and safety subgroup analyses by TE ADA categories may be performed.

#### 5.13.2.12. 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 listing of worsening follow-up dilated fundoscopic exam results will be generated with baseline and postbasline results.

The cases with repeated fundoscopy during the course of the trial, based on clinical suspicion of worsening retinopathy that have either findings of de novo retinopathy or progression of retinopathy, and severe/serious adverse events from PTs defined in searching criteria in Appendix 1 will be considered as AESIs and summarized.

### 5.13.2.13. Hepatobiliary Safety

#### *5.13.2.13.1. Hepatobiliary Disorders*

The AE database will be searched using SMQs to identify events consistent with hepatobiliary disorders. Detailed searching criteria for these AEs can be found in Appendix 1. A summary by treatment and PT within SMQ will be provided. Severe/serious hepatobiliary disorders will be considered AESIs.

#### *5.13.2.13.2. Acute Gallbladder Disease*

The AE database will be searched using pre-defined SMQs to identify events consistent with acute gallbladder diseases. Detailed searching criteria for these AEs can be found in Appendix 1. A summary by treatment and PT within SMQ will be provided. Severe/serious acute gallbladder diseases will be considered AESIs.

#### *5.13.2.13.3. Liver Enzymes*

Analyses for laboratory analyte measurements are described in Section 5.16. This section describes additional analyses of liver enzymes. In addition, the following will be provided by treatment groups:

- Shift table of maximum to maximum alanine aminotransferase (ALT) measurement from baseline (≤1×ULN, >1×ULN) to post-baseline with the following categories: ≤1×ULN, >1 to <3×ULN, ≥3 to <5×ULN, ≥5 to <10×ULN, ≥10×ULN.

- Shift table of maximum to maximum aspartate transaminase (AST) measurement from baseline (≤1×ULN, >1×ULN) to post-baseline with the following categories: ≤1×ULN, >1 to <3×ULN, ≥3 to <5×ULN, ≥5 to <10×ULN, ≥10×ULN.

- Shift tables of maximum to maximum total bilirubin and direct bilirubin from baseline to post-baseline with the following categories: ≤1×ULN, >1 to <2×ULN, ≥2×ULN.

- Shift tables of serum alkaline phosphatase (ALP) from baseline to post-baseline with the following categories: ≤1×ULN, >1 to 21×ULN, ≥2×ULN.

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

### 5.13.2.14. Gastrointestinal Safety

The time courses of prevalence and incidence (newly-occurring episodes) of nausea, vomiting, diarrhea, and combined will be plotted by treatment and maximum severity.

The maximum severity and duration of treatment-emergent nausea, vomiting, diarrhea, and combined through the end of the study will be summarized by treatment.

The PTs in the gastrointestinal SOC will be used to identify gastrointestinal AEs. The incidence of the resulting TEAEs will be summarized by treatment and PT. Preferred terms with serious/severe cases in the gastrointestinal SOC will be considered as AESIs.

#### 5.13.2.15. Acute Renal Events

Laboratory measures related to renal safety will be analyzed as specified for laboratory measurements in Section 5.16.

Additionally, two shift tables examining renal function will be created. A min-to-min shift table of estimated glomerular filtration rate (eGFR) estimated by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation with units ml/min/1.73m^2, using categories (<30, ≥30 to <45, ≥45 to <60, ≥60 to <90, and ≥90 mL/min/1.73m2). A max-to-max shift table of urine albumin-to-creatinine ratio (UACR), using the categories UACR<30 mg/g, 30 mg/g ≤ UACR ≤300 mg/g, UACR>300 mg/g (respectively, these represent normal, microalbuminuria, and macroalbuminuria).

The AE database will be searched using SMQs of acute renal failure and chronic kidney disease to identify events consistent with acute renal events. The incidence of the resulting TEAEs will be summarized by treatment and PT. Detailed searching criteria can be found in Appendix 1. Severe/serious acute renal events will be considered as AESI.

#### 5.13.2.16.Dehydration

The AE database will be searched using an SMQ of dehydration to identify events consistent with dehydration. Detailed searching criteria can be found in Appendix 1. Severe/serious dehydration events will be considered AESIs.

#### 5.13.2.17.Metabolic Acidosis, Including Diabetic Ketoacidosis

Adverse event database will be searched using MedDRA PT terms to identify events consistent with metabolic acidosis. Detailed searching criteria can be found in Appendix 1. Incidence of resulting TEAEs will be summarized by treatment and PT. Severe/serious metabolic acidosis, including diabetic ketoacidosis, will be considered an AESI.

#### 5.13.2.18.Amputation/Peripheral Revascularization

Adverse event database will be searched using MedDRA PT terms to identify events consistent with amputation or peripheral revascularization. Incidence of resulting TEAEs will be summarized by treatment and PT. Amputation/peripheral revascularization will be considered an AESI.

#### 5.13.2.19.Major Depressive Disorder/Suicidal Ideation

Adverse event database will be searched using MedDRA PT terms to identify events consistent with major depressive disorder or suicidal ideation. Detailed searching criteria can be found in Appendix 1. Incidence of resulting TEAEs will be summarized by treatment and PT.

Severe/serious major depressive disorder/suicidal ideation or behavior will be considered an AESI.

#### 5.13.2.20. Treatment of Overdose

Listing of patients reporting AEs related to over dosing of tirzepatide will be provided as a protocol deviation.

### *5.13.3. Renal Safety*

A subpopulation of "Patients with T2DM and renal dysfunction" is of study interest, will be identified as eGFR<60 mL/min/1.73m$^2$ or UACR >300 mg/g at baseline. The following analyses will be done in this subpopulation:

Summary and change from baseline in eGFR will be analyzed in MMRM analysis using actual measurements. The same analysis will be done for UACR in log scale. The analysis will be conducted utilizing for changes from baseline through post-baseline with a MMRM using restricted maximum likelihood (REML). The model will include stratification factors (country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), baseline SGLT2 inhibitor use (Yes or No)), treatment group, visit, and treatment-by-visit interaction as fixed effects, and baseline value of the corresponding renal marker (log scale if the corresponding renal parameter is analyzed on the log scale) as a covariate.

Shift table for eGFR and UACR will be done in the renal dysfunction subpopulation. Treatment-emergent adverse events in the renal dysfunction subpopulation will be summarized by treatment using MedDRA PT nested within SOC. Serious adverse events in the renal dysfunction subpopulation will be provided. Analyses of changes from minimum baseline value to minimum postbaseline value will be performed for eGFR. Analyses of changes from maximum baseline value to maximum postbaseline value will be performed for UACR, Calcitonin, ECG parameters and vitals (pulse, SBP and DBP). Severe hypoglycemia and level 2 hypoglycemia events will be summarized by treatment, excluding hypoglycemia events occurring after initiation of rescue therapy. Post-baseline threshold analysis will be done for eGFR (<15, <30, <45, and <60) and "eGFR <60 or UACR>300".

## 5.14. Vital Signs

Descriptive summaries by treatment and by nominal visits will be provided for the baseline and post-baseline values and change from baseline values. Two records taken at the same visit will be averaged first before being used for summarizing data and for analysis.

The MMRM using REML will be used to fit change from baseline in vital signs at all scheduled post-baseline visits. The model will include country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), baseline SGLT2 inhibitor use (Yes or No), treatment group, visit, and treatment-by-visit interaction as fixed effects, and baseline value of the dependent variable as a covariate. To model the covariance structure within patients, the unstructured covariance matrix will be used.

Counts and percentages of patients with abnormal sitting systolic blood pressure (BP), sitting diastolic BP, and pulse will be presented by treatment. The criteria for identifying patients with treatment-emergent vital signs abnormalities are stated in Table GPGM.5.2.

**Table GPGM.5.2. Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurement, and Categorical Criteria for Weight and Temperature Changes for Adults**

| Parameter | Low mmHg | High mmHg |
|---|---|---|
| Systolic BP (mm Hg) (Supine or sitting – forearm at heart level) | ≤90 and decrease from baseline ≥20 | ≥140 and increase from baseline ≥20 |
| Diastolic BP (mm Hg) (Supine or sitting – forearm at heart level) | ≤50 and decrease from baseline ≥10 | ≥90 and increase from baseline ≥10 |
| Pulse (bpm) (Supine or sitting) | <50 and decrease from baseline ≥15 | >100 and increase from baseline ≥15 |

## 5.15. Electrocardiogram

Summary statistics by treatment and by nominal visit will be provide for electrocardiogram (ECG) parameters (heart rate, PR, QRS, QT, and corrected QT using Fridericia's correction factor [QTcF]. When the QRS is prolonged (for example, a complete bundle branch block), QT and QTc should not be used to assess ventricular repolarization. Thus, for a particular ECG, the following will be set to missing (for analysis purposes) when QRS is ≥120: QT and QTcF.

The criteria for identifying patients with treatment-emergent quantitative ECG abnormalities are based on Table GPGM.5.3.

In addition, the percentages of patients with QT greater than 500 msec will be summarized, and the percentages of patients with QTcF greater than 500 msec will be summarized.

The percentages of patients who experienced a treatment-emergent increase from the maximum baseline in QTcF interval of greater than 30 msec, 60 msec, or 75 msec at any time will be summarized. Maximum baseline will be the maximum non-missing observation in the baseline period. The maximum value during the study follow up will be analyzed. Planned and unplanned measurements will be included.

**Table GPGM.5.3. Selected Categorical Limits for ECG Data**

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

## 5.16. Clinical Laboratory Evaluation

All laboratory data will be reported in International System of Units and Conventional Units. Out of reference range values will be flagged as high (H) or low (L) in the listings. Descriptive summaries by treatment and by nominal visits will be provided for the baseline and post-baseline values and change from baseline values for selected measurements.

Observed values and change from baseline values for selected measurements for each visit will be displayed graphically for patients who have both a baseline and a postbaseline planned measurement. Unplanned measurements will be excluded from graphs.

A shift table from baseline to postbaseline with unplanned measurements included will be provided for selected measurements. The shift table will include the number and percentage of patients within each baseline category (low, normal, high, or missing) versus each postbaseline category (low, normal, high, or missing) by treatment. Proportion of patients shifted will be compared between treatments using Fisher's exact test.

For qualitative laboratory analytes, the numbers and percentages of patients with normal and abnormal values will be summarized by treatment.

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

## 5.17. Health Outcomes

The patient-reported outcome questionnaires will be completed by the patients at baseline and at 52 weeks (or early termination visit prior to 52 weeks). These include use of the mITT population on the EAS, and use of a 2-sided alpha level of 0.05 and a 2-sided 95% CI for pairwise comparison. No multiplicity adjustment will be made in the evaluation of health

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

### *5.17.1. EQ-5D-5L*

Each item will be summarized descriptively by treatment at each scheduled visit at which the 5 level European Quality of Life – 5 dimensions (EQ-5D-5L) is administered. The changes from baseline to week 52 (last onvervation carried forward [LOCF]) in the index and visual analog scale (VAS) scores will be analyzed using an ANCOVA model with model terms of treatment, country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), and baseline SGLT2 inhibitor use (Yes or No) as fixed effects, and baseline EQ-5D-5L score as a covariate.

### *5.17.2. Impact of Weight on Self-Perceptions Questionnaire (IW-SP)*

Descriptive summaries by treatment at each scheduled visit at which the Impact of Weight on Self-Perceptions Questionnaire (IW-SP) is administered will be presented for each item. Treatment comparison in the raw and transformed overall IW-SP score change from baseline to week 52 (LOCF) will be analyzed using an ANCOVA model with model terms of treatment, country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), and baseline SGLT2 inhibitor use (Yes or No) as fixed effects, and baseline IW-SP score as a covariate.

### *5.17.3. Ability to Perform Physical Activities of Daily Living (APPADL)*

Descriptive summaries by treatment at each scheduled visit at which the Ability to Perform Physical Activities of Daily Living (APPADL) is administered will be presented for each item. Treatment comparison in the raw and transformed overall APPADL score change from baseline to week 52 will be analyzed using an ANCOVA model with model terms of treatment, country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), and baseline SGLT2 inhibitor use (Yes or No) as fixed effects, and baseline APPADL score as a covariate.

### *5.17.4. Diabetes Treatment Satisfaction Questionnaire (DTSQ)*

The Diabetes Treatment Satisfaction Questionnaire (DTSQ) contains 8 items (conceptually the same items in the status [DTSQs] and change [DTSQc] versions). Six items (1, and 4 through 8) are summed to produce a measure of treatment satisfaction and the two remaining items (2 and 3) are treated individually to assess, respectively, the perceived frequency of hyperglycemia and hypoglycemia. The DTSQs is used to assess treatment satisfaction at baseline and the DTSQc is used to assess relative change in satisfaction from baseline at week 52 or early termination.

Descriptive summaries will be provided at baseline (DTSQs only) and at 52 weeks (DTSQc only) for the perceived hyperglycemia item, perceived hypoglycemia item, the six-item overall satisfaction score.

Treatment comparison in the DTSQc at week 52 will be analyzed using an ANCOVA model with model terms of treatment, country/pooled country, baseline HbA1c (≤8.5%, >8.5% [≤69, >69 mmol/mol]), and baseline SGLT2 inhibitor use (Yes or No) as fixed effects and baseline DTSQs score as a covariate. The analyses will be conducted for the perceived hyperglycemia item, perceived hypoglycemia item, the 6-item overall satisfaction score.

## 5.18. Subgroup Analyses

Efficacy subgroup analyses will be guided by the efficacy estimand. HbA1c subgroup analysis will also be conducted relative to treatment-regimen estimand. Subgroup analyses may be done by country to support local regulatory registrations. Subgroup analysis will only involve clinically meaningful subgroups with adequate number of patients.

### *5.18.1. Subgroup Analysis of HbA1c Change at 52 Weeks*

Subgroup analyses by the following baseline characteristics will be provided: age group (<65, ≥65 years), age group (<75, ≥75 years), race, gender, ethnicity, region of enrollment (US, OUS), duration of diabetes (<median, ≥median), baseline HbA1c (≤8.5%, >8.5%), type of antihyperglycemic medication use (metformin alone, metformin+SU, metformin+SGLT2i, metformin+SU+SGLT2i, other), renal impairment (eGFR<60, ≥60), body mass index (BMI) group (<27, ≥27), and BMI group (<30, ≥30 to <35, ≥35).

### *5.18.2. Subgroup Analysis of Weight Change at 52 Weeks*

Subgroup analyses by the following baseline characteristics will be provided: age group (<65, ≥65 years), age group (<75, ≥75 years), race, gender, ethnicity, region of enrollment (US, OUS), duration of diabetes (<median, ≥median), baseline HbA1c (≤8.5%, >8.5%), type of antihyperglycemic medication use, renal impairment (eGFR<60, ≥60), BMI group (<27, ≥27), and BMI group (<30, ≥30 to <35, ≥35).

### *5.18.3. Subgroup Analysis of TEAE through safety follow up*

Subgroup analyses by the following baseline characteristics will be provided: age group (<65, ≥65 years), age group (<75, ≥75 years), race, gender, ethnicity, renal impairment (eGFR<60, ≥60), BMI group (<27, ≥27), and BMI group (<30, ≥30 to <35, ≥35).

Other exploratory subgroup analyses may be performed as deemed appropriate.

## 5.19. Interim Analyses and Data Monitoring Committee

A data monitoring committee (DMC) will have the responsibility for periodic review of unblinded interim analysis results in order to monitor the safety of the patients in the study. A statistical analysis group external to the study team will perform the data analysis for the DMC. No interim analyses of efficacy are conducted with a view of early study termination or study modification. Thus interim analyses will have no bearing on type 1 error controls associated with final efficacy analyses. Detailed information regarding safety reviews including the statistical reports to be reviewed by the DMC will be described in the program DMC Charter.

### *5.19.1. Interim Safety CV Meta-Analysis*

This study is designed to contribute the majority MACE-4 endpoints for the tirzepatide program-wide meta-analysis intended to discharge excessive cardiovascular risk with tirzepatide. The analyses of these MACE-4 endpoints will demonstrate that tirzepatide treatment is not associated with excessive CV risk. There is a planned interim analysis of this safety CV meta-analysis. An interim data extraction from this study will be conducted to support the interim safety CV meta-

analysis. The interim safety CV meta-analysis will have no bearing on type 1 error controls associated with efficacy analyses for this study. For more details please refer to the study procotol and CV meta-analysis statistical analysis plan.

## 5.20. COVID-19 Impact Assessment

This section lists the potential statistical analyses that may be performed to assess the impact of COVID-19 pandemic when appropriate.

### *5.20.1. Patients Impacted by COVID-19*

Listings of patients with protocol deviation or mitigation due to COVID-19, patients with COVID-19 AEs or death, and patients dispositions with reasons related to COVID-19 will be provided.

### *5.20.2. Adverse Events*

A summary table for patients with AEs related to COVID-19, including death due to COVID-19, serious COVID-19 AEs, and COVID-19 AEs, will be provided by study treatment.

### *5.20.3. Patient Disposition*

Patient disposition with reasons related to COVID-19 (such as COVID-19 AE, patient decision, etc.) will be summarized for study and study treatment discontinuation by treatment group.

### *5.20.4. Study Visits*

A summary of patients with study visit impacted by COVID-19 will be provided by treatment group. In this table, number and proportion of patients missing study visit including primary endpoint visit, having home health visit and virtual visit will be summarized.

### *5.20.5. Mitigation Summary*

A summary table for patients having protocol deviation and mitigation due to COVID-19 (such as missing study visit, having home health visit, etc.) will be provided by treatment group. An additional summary may be provided by country of enrollment and treatment group.

### *5.20.6. Measures Related to Primary and Key Secondary Objectives*

Patients missing measures (HbA1C, fasting glucose, and body weight) related to primary and key secondary objectives will be summarized by visit and treatment group. In addition, the number of patients utilizing alternative options to in-person visits (such as local lab, home health visits, etc.) to collect primary and key secondary measures may be summarized by visit and treatment group.

# 6. Unblinding Plan

Please refer to the I8F-MC-GPGM Blinding and Unblinding Plan.

# 7. References

Protocol I8F-MC-GPGM (a) Efficacy and Safety of LY3298176 Once Weekly versus Insulin Glargine in Patients with Type 2 Diabetes and Increased Cardiovascular Risk (SURPASS-4).

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

# Appendix 1.Searching Criteria for Adverse Events of Special Interest

The AESI analyses are detailed in Section 5.13.2. The search criteria for each AESI are stored in CLUWE with path: \\statsclstr\lillyce\prd\ly3298176\common\AESI_Lab\Search criteria AESIs_TZP.xlsx.

**Leo Document ID = b46cbdc0-65ea-4dbb-aeda-dbdf4a4375e8**

**Approver:** PPD

**Approval Date & Time: 15-Feb-2021 19:34:54 GMT**

**Signature meaning: Approved**