Statistical Analysis Plan (V2): I8F-MC-GPGI

A Randomized, Phase 3, Double-blind Trial Comparing the Effect of the Addition of Tirzepatide Versus Placebo in Patients With Type 2 Diabetes Inadequately Controlled on Insulin Glargine With or Without Metformin

NCT04039503

Approval Date: 01-Feb-2021

1. Statistical Analysis Plan: I8F-MC-GPGI: A Randomized, Phase 3, Double-blind Trial Comparing the Effect of the Addition of Tirzepatide versus Placebo in Patients with Type 2 Diabetes Inadequately Controlled on Insulin Glargine with or without Metformin (SURPASS-5)

Tirzepatide (LY3298176) Type 2 Diabetes Mellitus

A phase 3, randomized, double-blind trial comparing the effect of the addition of tirzepatide to placebo in patients with type 2 diabetes inadequately controlled on Insulin Glargine with or without metformin.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I8F-MC-GPGI
Phase 3

Statistical Analysis Plan electronically signed and approved by Lilly on 20 August 2020. Statistical Analysis Plan v2 electronically signed and approved by Lilly date provided below.

Approval Date: 01-Feb-2021 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Table of Contents

3. Revision History

Statistical analysis plan (SAP) Version 1 was approved prior to the first unblinded data transfer.

The second version is approved before the final database lock. The following represent major changes made for the second version:

1. Updated definition of analysis set. Per agreement with the US FDA, exclude patients discontinuing study drug due to inadvertent enrollment from efficacy analyses
2. Updated baseline definition for selected measures.
3. Missing data imputation: Modified the definition of "retrieved dropouts"
4. Added language to allow for the use of local laboratory data when central laboratory data are not available for glycemic control measures
5. Included proportion of patients achieving HbA1c <5.7% as a key secondary endpoint subject to type 1 error rate control for TZP 15 mg and TZP 10 mg versus placebo
6. 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
7. Added section to assess SARS-CoV-2 (COVID-19) impact

4. Study Objectives

4.1. Primary Objectives

The primary objectives of the study are to demonstrate superiority of once-weekly (QW) tirzepatide 10 mg and/or 15 mg versus placebo when added to titrated basal insulin glargine, with or without metformin, with respect to mean change in hemoglobin A1c (HbA1c) from baseline at 40 weeks.

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

Together with the primary objectives, the following secondary objectives are subject to strong control of the type 1 error rate (see Section 6.12.3):

- To demonstrate superiority of QW tirzepatide 5 mg versus placebo when added to titrated basal insulin glargine, with or without metformin, with respect to mean change in HbA1c from baseline at 40 weeks.
- To demonstrate superiority of QW tirzepatide 5 mg, 10 mg, and/or 15 mg versus placebo when added to titrated basal insulin glargine with or without metformin at 40 weeks for
    - mean change in body weight from baseline
    - the proportion of patients with HbA1c target values of <7.0% (53 mmol/mol), and
    - mean change in fasting serum glucose (central laboratory) from baseline.
- To demonstrate superiority of QW tirzepatide 10 mg, and/or 15 mg versus placebo when added to titrated basal insulin glargine with or without metformin at 40 weeks for
    - the proportion of patients with HbA1c target values of <5.7% (39 mmol/mol)

4.3. Other Secondary and Exploratory Efficacy Objectives Not Subject to Type 1 Error Rate Control

The following secondary efficacy objective is not subject to strong control of the type 1 error rate:

- To compare QW tirzepatide 5 mg, 10 mg, and 15 mg to placebo at 40 weeks for
    - the proportion of patients achieving HbA1c target ≤6.5% (48 mmol/mol)
    - mean change in daily average 7-point self-monitored blood glucose profiles from baseline
    - the proportion of patients achieving weight loss of ≥5%, ≥10%, and ≥15% from baseline; and
    - the change from baseline in daily mean insulin glargine dose.
- To compare QW tirzepatide 5 mg to placebo at 40 weeks for

- the proportion of patients achieving HbA1c target <5.7% (39 mmol/mol)

The following tertiary/exploratory objective is also not subject to strong control of the type 1 error rate:

- To compare QW tirzepatide 5 mg, 10 mg, and 15 mg to placebo at 40 weeks for:
  - mean change in lipids (total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), very low-density lipoprotein (VLDL), and triglycerides (TG)
  - mean change in waist circumference
  - changes from baseline in mean body mass index (BMI)
  - biomarkers, and
  - patient-reported outcomes:
    - Ability to Perform Physical Activities of Daily Living (APPADL)
    - Impact of Weight on Self-Perception (IW-SP)
    - Diabetes Treatment Satisfaction Questionnaire status (DTSQs)/Diabetes Treatment Satisfaction Questionnaire change (DTSQc), and
    - EQ-5D-5L.

4.4. Safety Objectives

The following safety objectives are as follows:

- To compare QW tirzepatide 5 mg, 10 mg, and 15 mg to placebo through the end of the safety follow-up period for
  - treatment-emergent adverse events (TEAEs)
  - early discontinuation of study drug due to adverse events (AEs)
  - adjudicated pancreatic AEs
  - serum calcitonin
  - incidence of allergic and hypersensitivity reactions
  - incidence of treatment-emergent antidrug antibodies to tirzepatide
  - mean change in systolic and diastolic blood pressure and heart rate from baseline
  - occurrence of hypoglycemic episodes, and
  - incidence of initiation of rescue therapy for severe, persistent hyperglycemia.

4.5. Pharmacokinetics

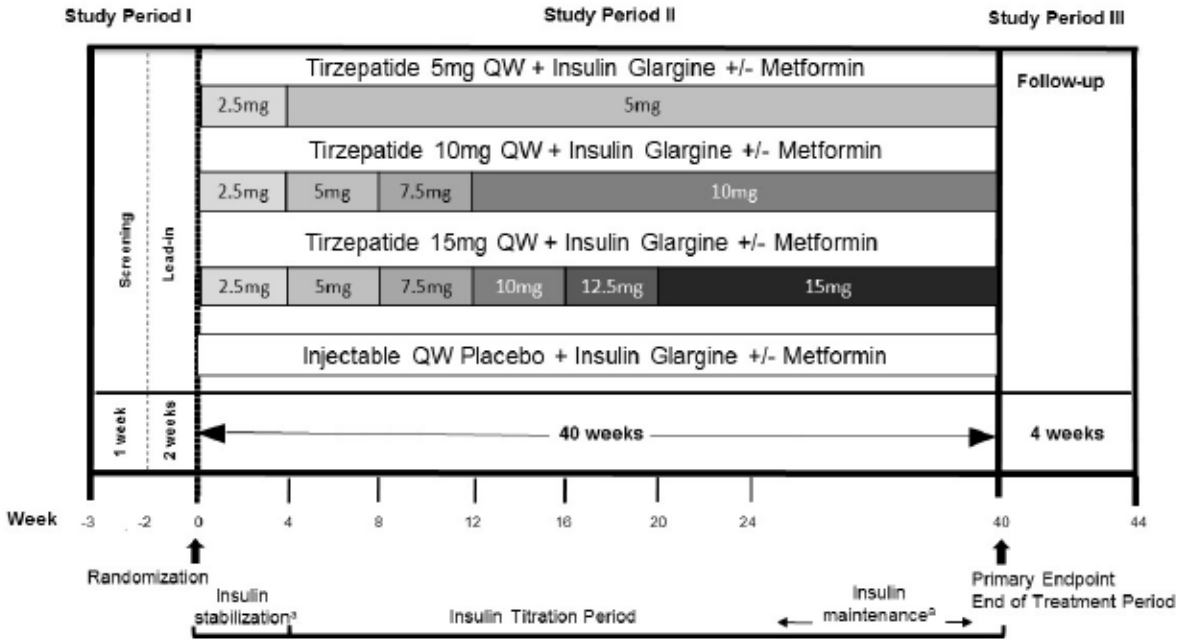
To characterize the pharmacokinetics (PK) of QW tirzepatide 5 mg, 10 mg, and 15 mg doses and evaluate the relationship between tirzepatide exposure and safety, tolerability, and efficacy measures for population PK and PD parameters.

5. Study Design

5.1. Summary of Study Design

Study I8F-MC-GPGI (GPGI) is a multicenter, randomized, double-blind, parallel, multinational, placebo-controlled Phase 3 study which will assess the safety and efficacy of the addition of 5 mg, 10 mg, or 15 mg tirzepatide, or placebo for change from baseline in HbA1c in patients with type 2 diabetes mellitus (T2DM) receiving titrated basal insulin glargine (with or without metformin) over 40 weeks of treatment.

Figure GPGI.5.1 illustrates the study design.



Abbreviation: QW = once weekly.

[a] Stabilization Period = first 4 weeks after randomization, with restricted insulin dose adjustments. Insulin Glargine Titration Period Weeks 4 to 40 (end of treatment/end of study), with unrestricted insulin dose adjustments. Maintenance Period = Weeks 24 to 40 (end of treatment/end of study), the period when insulin glargine dose is expected to be stable.

Figure GPGI.5.1. Illustration of study design for Clinical Protocol I8F-MC-GPGI.

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. A dilated fundoscopic exam will be performed between Visit 2 and Visit 3 as results are required to confirm eligibility.

Study Period II (40-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.

Treatment Period: General Considerations

The treatment period will last 40 weeks, starting with a 4-week stabilization period immediately after randomization and followed by a 36-week glargine titration period. The maintenance period is defined as a part of the titration period when the insulin glargine dose is expected to be stable and optimized (Weeks 24 to 40).

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 in the 5 mg group. For the 10 mg group, 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 group, 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 to 15 mg) until the 15 mg dose is reached and maintained for the duration of the study. For the placebo group, patients will inject matched QW placebo for the duration of the study.

Postrandomization period (end of Visit 3 to Visit 14)

Stabilization Period (End of Visit 3 through Visit 7)

The main purposes of this period are to introduce randomized study drugs (QW tirzepatide or QW placebo) in a safe manner and to assure regular and correct use of the self-monitoring and insulin dose adjustment procedures and study diaries during the entire study. In an effort to allow appropriate time for tirzepatide to reach steady state, insulin glargine dose adjustments during the 4-week stabilization period should be restricted to those needed in case of significant safety risks due to inadequate insulin dose (hypoglycemia or severe hyperglycemia) in which case patients should be instructed to contact the sites to adjust the insulin glargine dose per the treat-to-target (TTT) algorithm. In addition, for patients with baseline HbA1c $\leq$8.0%, the insulin glargine dose will be decreased by 20% immediately after randomization, no later than 7 days after the first dose of study drug and will then remain unchanged during the stabilization period to decrease the

risk of hypoglycemia. The insulin glargine dose will remain unchanged if baseline HbA1c is >8.0%.

Titration Period (End of Visit 7 through Visit 22)

At the beginning of the titration period, the patients will be instructed to start using the TTT algorithm without restrictions in order to reach the optimal dose of insulin glargine as soon as possible. The patients will be requested to perform an insulin dose assessment once weekly during this period.

Study Period III (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 last visit. During the safety follow-up period, patients will not receive study drug. Initiation of new antihyperglycemic therapy for the safety follow-up period will not be classified as "rescue therapy."

5.2. Sample Size Determination

The trial is powered to assess superiority of tirzepatide 10 mg and 15 mg relative to the primary endpoint (mean change in HbA1c from baseline to 40 weeks).

The power is assessed based on the following assumptions:

- each of the 10- and 15-mg tirzepatide treatment groups will be tested in parallel against placebo at a 2-sided 0.025 significance level
- use of a 2-sample t-test utilizing HbA1c data collected before initiation of any rescue medication or premature treatment discontinuation with no more than 28% of subjects initiating rescue medication or prematurely discontinuing treatment in each treatment group
- 0.6% greater mean reduction in HbA1c from baseline for 10 and 15 mg tirzepatide compared with placebo
- 1:1:1:1 randomization, and
- a common standard deviation (SD) of 1.1%.

On the basis of these assumptions, a sample size of 472 subjects is required to ensure at least 90% power to demonstrate that tirzepatide 10 mg and/or 15 mg are superior to placebo relative to the primary endpoint. Furthermore, this sample size will ensure 90% power for the superiority evaluation conducted using an analysis of covariance (ANCOVA) utilizing all available HbA1c data at 40 weeks with missing data imputed with a conservative multiple imputation method (as described in Section 6.12.1.3), provided a 0.6% greater mean reduction in HbA1c from baseline for 10 and 15 mg tirzepatide compared with placebo and SD increases to no more than 1.3% due

to the inclusion of data on rescue medications and after premature treatment discontinuation and imputation of missing data.

5.3. Method of Assignment to Treatment

Approximately 472 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 (IWRS). Patients will be randomized in a 1:1:1:1 ratio to receive 5 mg tirzepatide, 10 mg tirzepatide, 15 mg tirzepatide, or placebo. The randomization will be stratified by country, baseline HbA1c concentration (≤8.0%, >8.0% [≤64, >64 mmol/mol]), and baseline metformin use (Yes/No).

6. A Priori Statistical Methods

6.1. Populations for Analyses

For purposes of analysis, Table GPGI.6.1 defines analysis populations/data sets.

Table GPGI.6.1. 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 intention-to-treat (mITT) population	All randomly assigned participants who are exposed to at least 1 dose of study drug. In the event of a treatment error, participants will be analyzed according to the treatment they were randomized.
Efficacy analysis set (EAS)	Data obtained during Study Period II from the mITT population, excluding patients discontinuing study drug due to inadvertent enrollment and data after initiating rescue antihyperglycemic medication or early discontinuation of study drug (last dose date + 7 days).
Full analysis set (FAS)	All available data obtained during Study Period II from the mITT population, regardless of adherence to study drug or initiation of rescue antihyperglycemic medication. Patients who discontinued study drug due to inadvertent enrollment will be excluded.
Safety analysis set (SS)	All available data obtained during Study Periods II and III from the mITT population, regardless of adherence to study drug or initiation of new antihyperglycemic medication.

6.2. 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 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). Listings 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).

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.

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 lab and ECG, baseline needs to be prior to or within one hour after the first dose time.
- For patient-reported outcomes, data collected at Visit 3, regardless of the timing relative to first dose, will serve as baseline.

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

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

The 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 the 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 the 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 analyses. A 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, SD, median, minimum, and maximum. The summary statistics will be presented by nominal visit.

Statistical treatment comparisons will only be performed between tirzepatide doses and placebo. Comparisons among tirzepatide arms will not be performed unless specified otherwise.

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

6.3. Adjustments for Covariates

The study is stratified by country, HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]) at screening, and baseline metformin use (Yes/No). Where necessary to be included as a fixed effect in statistical models, countries with fewer than 10 randomized patients will be pooled into one category (pooled country). For HbA1c related analyses, country/pooled country and baseline metformin use (Yes/No) will be used as fixed effects and baseline HbA1c as a covariate. For

other efficacy analyses, country/pooled country, baseline HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]), and baseline metformin use (Yes/No) will be used as fixed effects and respective baseline value as a covariate. Fixed effects will be based on the information collected in the clinical database.

6.4. Handling of Dropouts or Missing Data

For the primary and secondary efficacy endpoint analyses subject to type 1 error rate control, data for patients with missing values at the 40-week visit will be imputed based on the method described in Section 6.12.1.3. Unless specified otherwise, imputation of missing data will be limited to primary and key secondary efficacy endpoint analyses. Any other secondary or exploratory efficacy parameter values or safety laboratory values that are missing will not be explicitly imputed.

6.5. Multicenter Studies

To investigate potential regional influences on efficacy, country/pooled country will be used as a fixed effect in the primary and secondary efficacy analyses.

6.6. Multiple Comparisons/Multiplicity

The type 1 error rate control strategy for the primary and key secondary efficacy objectives is illustrated in Section 6.12.3. No multiplicity adjustments will be made for conducting separate analyses relative to the "efficacy" and "treatment-regimen" estimands, evaluating other secondary or exploratory efficacy objectives, or safety assessments.

6.7. Patient Disposition

Reasons for screen failures 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 the study treatment disposition for all randomized patients will be summarized by planned study treatment.

6.8. Patient Characteristics

A listing of patient demographics will be provided for all randomized patients. All demographic and baseline clinical characteristics will be summarized by study treatment for all randomized patients. Baseline demographic and clinical characteristics of special interest include

- age
- gender
- race
- ethnicity
- weight
- country of enrollment

- HbA1c
- fasting serum glucose
- insulin glargine daily dose, and
- duration of T2DM.

6.9. Prior and Concomitant Therapy

Summary of prior glucagon-like peptide-1 receptor agonist (GLP-1 RA) use by type and treatment arm will be provided.

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

Concomitant medication summaries of interest include

- baseline antihypertensive therapy, by type
- baseline lipid lowering therapy, by type
- baseline antihyperglycemic therapy
- utilization of new antihyperglycemic therapy during Study Period II, and also during combined Study Periods II and III
- changes to baseline medication in Period II:
    - antihypertensive therapy
    - lipid lowering therapy
    - antihyperglycemic therapy
- changes to baseline medication during combined Study Periods II and III:
    - antihyperglycemic therapy
- rescue therapy and
- initiation of following medication in Study Period II:
    - antidiarrheal medication
    - antiemetic medication.

6.10. Treatment Exposure and Compliance

A 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 (if the data is available).

A summary of duration to follow-up (defined as time in days from date of randomization to date of safety follow-up, date of early study discontinuation or date of death) 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 study treatment.

6.10.1. Exposure and Compliance

The number of patients prematurely discontinuing study treatment prior to the 40-week visit will be provided by study treatment. Reasons for prematurely discontinuing study treatment prior to the 40-week visit will be provided by study treatment. Time-to-event analysis of premature study treatment discontinuation will be conducted.

A listing of patients who re-initiate tirzepatide/placebo due to missing ≥3 consecutive doses may be produced if data warrants.

Summary information on total daily dose of insulin glargine will be reported by visit. Information related to insulin glargine in regard to compliance to treat-to-target algorithm including number of assessments performed correctly, number of assessments that required a dose change, number of assessments for which the outcomes were correctly followed by the patient, and reasons for noncompliance will be summarized.

Overall treatment compliance will be defined as taking at least 75% of the scheduled study drug or insulin glargine injections. Study drug compliance will be calculated by taking the number of injections (regardless of the actual dose administered) divided by the total number of injections expected to be administered × 100. Insulin glargine compliance will be calculated by taking the total 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. Overall treatment compliance will be summarized descriptively by treatment using the mITT population.

6.11. Important Protocol Deviations

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

6.12. Efficacy Analyses

For the Food and Drug Administration (FDA) and potentially for other regulatory agencies, primary and key secondary efficacy assessments will be guided by the "treatment-regimen" estimand conducted using the 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 6.12.1.3) at 40 weeks. For publications and other purposes, the

assessment of efficacy objectives will be guided by the "efficacy" estimand using the EAS without imputation of missing data. A listing of patients randomized but not included in efficacy analyses (not treated or discontinued treatment due to inadvertent enrollment) will be provided.

6.12.1. Primary Efficacy Analysis

The primary efficacy measure will be change in HbA1c (% and mmol/mol) from baseline (postbaseline – baseline). 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 the 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 the analysis.

6.12.1.1. Primary Analysis Relative to the Efficacy Estimand

The primary analysis relative to the "efficacy" estimand will be conducted using HbA1c data in the EAS from baseline through the 40-week visit with the aid of a mixed-model repeated measure (MMRM). Restricted maximum likelihood (REML) will be used to obtain model parameter estimates and the 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 of interest will include treatment, visit, treatment-by-visit interaction, country/pooled country, and baseline metformin use (Yes/No) as fixed effects, and baseline HbA1c as a covariate. An unstructured covariance matrix will be used to model the within-patient errors. If this model fails to converge, the following covariance structures will be tested in the following order:

1. Heterogeneous Toeplitz
2. Heterogeneous First Order Autoregressive
3. Heterogeneous Compound Symmetry
4. Toeplitz
5. First Order Autoregressive, and
6. Compound Symmetry.

The first covariance structure that converges will be used. The resulting least squares mean (LSM) estimate of mean change from baseline in HbA1c will be summarized by visit and by study treatment.

With the aid of the MMRM analysis, p-values, and 2-sided 95% confidence intervals (CIs) for mean change in HbA1c from baseline to the 40-week visit will be derived and summarized for the 5 mg, 10 mg, and 15 mg doses of tirzepatide compared to placebo.

6.12.1.2. Primary Analysis Relative to the Treatment-Regimen Estimand

The primary analysis relative to the treatment-regimen estimand will be conducted utilizing HbA1c data in the FAS at baseline and at the 40-week visit with the aid of an ANCOVA model.

The response variable will be the primary measure and model terms will include treatment, country/pooled country, and baseline metformin use (Yes/No) as fixed effects and baseline HbA1c as a covariate. The ANCOVA analysis will be conducted using multiple imputation of missing primary measures (see Section 6.12.1.3 for details) and statistical inference over multiple imputation of missing data guided by Rubin (1987).

With the aid of the ANCOVA analysis, p-values and 2-sided 95% CIs for mean change in HbA1c from baseline to the 40-week visit will be derived and summarized for the 5 mg, 10 mg, and 15 mg doses of tirzepatide compared to placebo.

6.12.1.3. Methods for Multiple Imputations

For efficacy analyses relative to the "treatment-regimen" estimand, missing HbA1c data at the 40-week visit will be imputed based on "retrieved dropouts," defined as patients who had their HbA1c value measured at the 40-week visit in the same treatment arm who prematurely discontinued study drug. A pseudo-SAS program for implementing multiple imputations using data from retrieved dropouts is included in Appendix 1. In cases where there are not enough retrieved dropouts to provide a reliable imputation model, an alternative multiple imputation method with reference to the placebo group (placebo multiple imputation) will be used as the primary analysis relative to the treatment-regimen estimand. Analyses will be conducted with multiple imputations, and statistical inference over multiple imputations will be guided by the method proposed by Rubin (1987). If the primary analysis using "retrieved dropouts" for imputation converges, the analysis using "placebo multiple imputation" to impute the missing data will be conducted as a sensitivity analysis. If value of the imputed 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.

6.12.2. Secondary Efficacy Analyses Subject to Type 1 Error Rate Control

6.12.2.1. Analyses of Continuous Outcomes

Analyses for change from baseline in body weight as well as fasting serum glucose (FSG; postbaseline – baseline) will be conducted in a manner similar to the primary analyses in Section 6.12.1. Baseline HbA1c category (≤8.0%, >8.0% [≤64, >64 mmol/mol]) will be used as a fixed factor in place of baseline HbA1c as a covariate and baseline of the corresponding variable will be used as an additional covariate in the statistical model. Least squares mean estimates of mean change in body weight and FSG from baseline will be summarized by nominal visit and by study treatment. When applicable, fasting glucose data from a local lab will be used when central lab data is not available. For the multiple imputation of missing values, if value of the imputed weight change from baseline is <-50 kg or >50 kg, that value will be set to -50 kg or 50 kg, respectively, to avoid unrealistic imputed values; if value of the imputed fasting serum glucose change from baseline is <-20 mmol/L or > 20 mmol/L, that value will be set to -20 mmol/L or 20 mmol/L, respectively, to avoid unrealistic imputed values.

6.12.2.2. Analyses of Binary Outcomes

The analysis relative to the "efficacy" estimand for the endpoint at 40 weeks will be performed using EAS with missing values imputed from an MMRM model and then dichotomized. The MMRM model includes treatment, country/pooled country, baseline metformin use (Yes/No), 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, baseline metformin use (Yes/No) as fixed effects, and baseline HbA1c as a covariate. In addition, analysis will be conducted utilizing data using EAS from baseline through the 40-week visit with the aid of a longitudinal logistic regression with repeated measurements with treatment, visit, treatment-by-visit interaction, country/pooled country, and baseline metformin use (Yes/No) as fixed effects and baseline HbA1c as a covariate. In case the longitudinal logistic model does not converge due to small number of events, logistic regression will be utilized to analyze proportion of patients achieving HbA1c targets at nominal visits.

The analysis relative to the "treatment-regimen" estimand will be conducted utilizing HbA1c data in the FAS at baseline and at the 40-week visit with the aid of a logistic regression using multiple imputation of missing HbA1c data at the 40-week visit (see Section 6.12.1.3 for details). Model terms will include treatment, country/pooled country, and baseline metformin use (Yes/No) as fixed effects, and baseline HbA1c as a covariate. Statistical inference over multiple imputations will be guided by Rubin (1987).

6.12.3. Type 1 Error Rate Control Strategy for Primary and Key Secondary Efficacy Analyses

Since they are intended for different purposes, no type 1 error rate adjustments will be made for conducting analyses relative to "efficacy" and "treatment-regimen" estimands. For analyses within each estimand, the type 1 error control strategy for evaluation of the primary and key secondary objectives is illustrated in Figure GPGI.6.1.

The type I error rate control strategy for the Pharmaceuticals and Medical Devices Agency (PMDA) is in Appendix 2.

1. $H_{15,1}$ and $H_{15,2}$ are evaluated hierarchically each at a 2-sided 0.025 significance level conditioned on successfully achieving the preceding objective. In parallel,
2. $H_{10,1}$ and $H_{10,2}$ are evaluated hierarchically each at a 2-sided 0.025 significance level conditioned on successfully achieving the preceding objective.
3.
    a. If all objectives in #1 and #2 above are successfully established, $H_{5,1}$ and $H_{5,2}$ are evaluated hierarchically, each at a 2-sided 0.05 significance level.
    b. If all objectives in only #1 or only #2 above are successfully established, $H_{5,1}$ and $H_{5,2}$ are evaluated hierarchically, each at a 2-sided 0.025 significance level.
4. If both objectives: $H_{5,1}$ and $H_{5,2}$ are successfully established and
    a. If all objectives in #1 and #2 above are successfully established, then $H_{10,3}$, $H_{15,3}$, $H_{10,4}$, $H_{15,4}$, $H_{5,3}$, $H_{5,4}$, $H_{15,5}$, and $H_{10,5}$ will be evaluated hierarchically each at a 2-sided 0.05 significance level conditioned on successfully achieving the preceding objective.
    b. If all objectives in only #1 or only #2 above are successfully established, then $H_{10,3}$, $H_{15,3}$, $H_{10,4}$, $H_{15,4}$, $H_{5,3}$, $H_{5,4}$, $H_{15,5}$, and $H_{10,5}$ will be evaluated hierarchically each at a 2-sided 0.025 significance level conditioned on successfully achieving the preceding objective.
5. If all objectives in #3, #4, and #5 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.

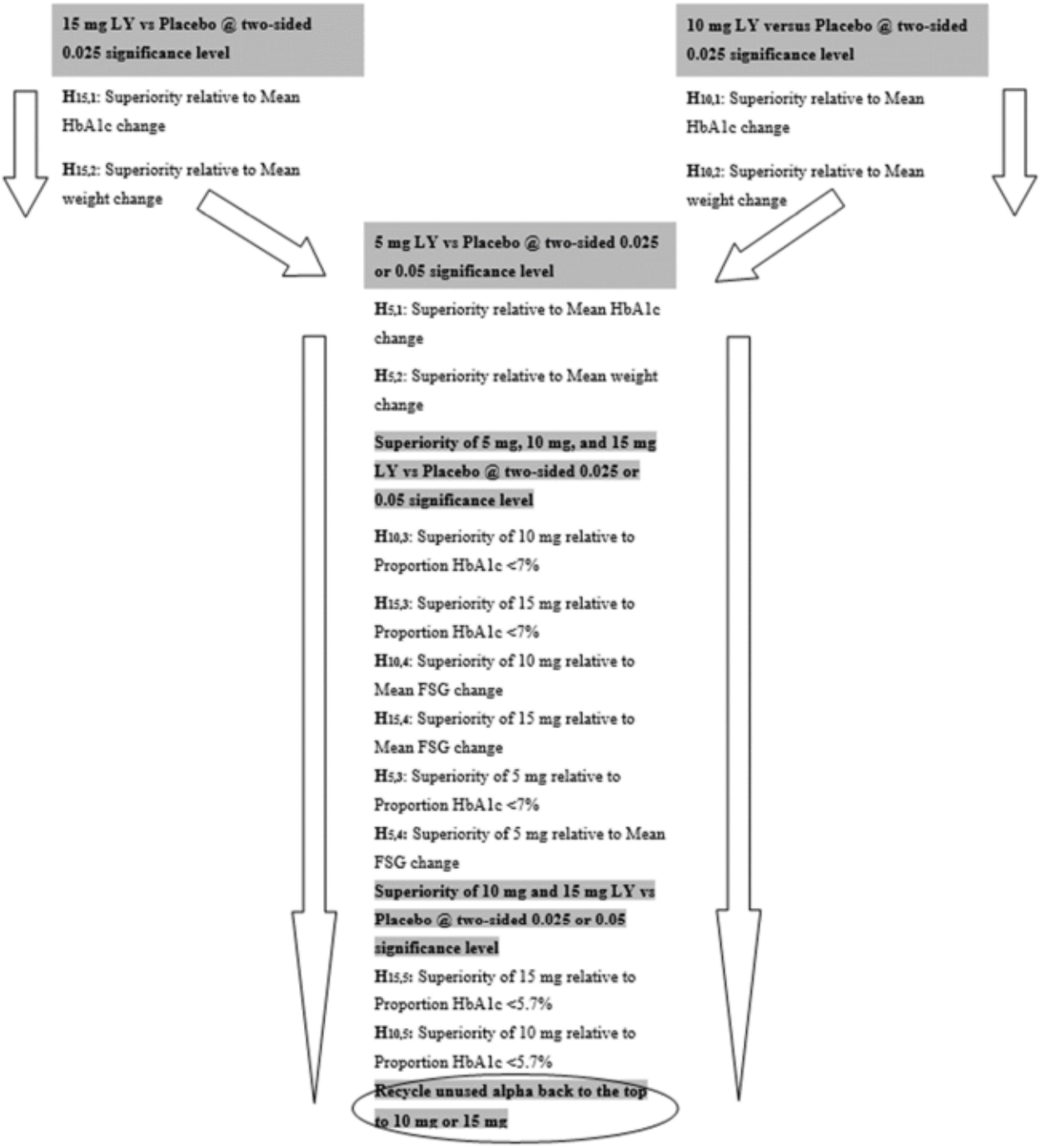


Figure GPGI.6.1. Type 1 error control strategy for primary and key secondary efficacy endpoints.

6.12.4. Other Secondary and Exploratory Efficacy Analyses

Other secondary and exploratory efficacy measures will use the efficacy estimand and will be summarized by treatment and nominal visit. Unless otherwise specified, missing data will not be

explicitly imputed, and assessments are not subject to type 1 error rate control. Some parameters may be log transformed, if necessary. Biomarkers analyses are to be determined.

Table GPGI.6.2. Secondary and Exploratory Efficacy Analyses Not Controlled for Type I Error

Objective	Relative to the efficacy measure:	Analysis Conducted in a manner similar to section:	Additional Information
Secondary Analyses			
To compare QW tirzepatide 5 mg, 10 mg, and 15 mg to placebo at 40 weeks in:	Proportion of patients achieving an HbA1c target value of ≤6.5% (48 mmol/mol)	6.12.2.2	None
	Change from baseline in 7-point self-monitored blood glucose (SMBG) profiles	6.12.1.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline SMBG parameter as a covariate. LSM estimates at 40-weeks will be summarized by treatment and 7-points. In addition to the analyses on each of the 7-points, similar analyses will be done for the 2-hour morning, midday, and evening meal excursions, the mean of all meals 2-hour excursion, the mean of all 7-point measurements, the mean of all pre-meal measurements, and the mean of all 2-hour postprandial measurements.
	Proportion of patients who achieved weight loss of ≥5%, from baseline	6.12.2.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline weight as a covariate

Secondary and Exploratory Efficacy Analyses Not Controlled for Type I Error

Objective	Relative to the efficacy measure:	Analysis Conducted in a manner similar to section:	Additional Information
	Proportion of patients who achieved weight loss of ≥10% from baseline	6.12.2.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline weight as a covariate
	Proportion of patients who achieved weight loss of ≥15% from baseline	6.12.2.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline weight as a covariate
	Change from baseline in daily mean insulin glargine dose	6.12.1.1	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline insulin glargine dose as a covariate.
To compare QW tirzepatide 5 mg to placebo at 40 weeks in:	Proportion of patients achieving an HbA1c target value of <5.7% (39 mmol/mol)	6.12.2.2	None
Exploratory Objectives			
To compare QW tirzepatide 5 mg, 10 mg, and 15 mg to placebo at 40 weeks in:	Change from baseline in lipid parameters (Total-Cholesterol, HDL-C, LDL-C, VLDL-C, TG)	6.12.1.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use corresponding baseline lipid parameter as a covariate
	Change from baseline in waist circumference	6.12.1.1	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline waist circumference as a covariate
	Change from baseline in BMI	6.12.1.1	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline BMI as a covariate
	Change from baseline in patient reported outcomes: APPADL, IW-SP, EQ-5D-5L	6.12.1.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use corresponding baseline patient outcome score as a covariate.

Secondary and Exploratory Efficacy Analyses Not Controlled for Type I Error

Objective	Relative to the efficacy measure:	Analysis Conducted in a manner similar to section:	Additional Information
	Patient reported DTSQc score	6.12.1.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use DTSQs baseline patient outcome score as a covariate
	Proportion of patients achieving an FSG <100 mg/dL (5.5 mmol/L)	6.12.2.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline FSG as a covariate.
	Proportion of patients achieving an FSG <126 mg/dL (7 mmol/L)	6.12.2.2	Use baseline HbA1c category as a fixed effect in place of baseline HbA1c as a covariate. Use baseline FSG as a covariate.
To compare QW tirzepatide 5 mg, 10 mg, and 15 mg to placebo at 40 weeks in:	Proportion of patients achieving an HbA1c target ≤6.5%, without weight gain (<0.1 kg), and without documented symptomatic hypoglycemia or severe hypoglycemia	6.12.2.2	Include baseline body weight as an additional covariate
	Proportion of patients achieving an HbA1c target <7.0% without weight gain (<0.1 kg), and without documented symptomatic hypoglycemia or severe hypoglycemia	6.12.2.2	Include baseline body weight as an additional covariate
	Proportion of patients achieving an HbA1c target ≤6.5% without weight gain (<0.1 kg), and without clinically significant documented symptomatic hypoglycemia or severe hypoglycemia	6.12.2.2	Include baseline body weight as an additional covariate
	Proportion of patients achieving an HbA1c target <7.0% without weight gain (<0.1 kg), and without clinically significant documented symptomatic hypoglycemia or severe hypoglycemia	6.12.2.2	Include baseline body weight as an additional covariate

Secondary and Exploratory Efficacy Analyses Not Controlled for Type I Error

Abbreviations: APPADL = Ability to Perform Physical Activities of Daily Living; BMI = body mass index; DTSQc = Diabetes Treatment Satisfaction Questionnaire change; DTSQs = Diabetes Treatment Satisfaction Questionnaire change; FSG = fasting serum glucose; HbA1c = hemoglobin A1c; HDL-C = high-density lipoprotein-cholesterol; IW-SP = Impact of Weight on Self-Perceptions Questionnaire; LDL-C = low-density lipoprotein cholesterol; LSM = least squares mean; QW = once-weekly; TG = triglyceride; VLDL-C = very-low density lipoprotein-cholesterol.

6.13. Safety Analyses

Unless specified otherwise, safety assessments will be based on the SS (see Table GPGI.6.1). All events that occur between the date of first dose of study drug to the date of the patient's safety follow-up visit or the patient's end of study participation will be included. For assessing the benefit risk profile through 40 weeks, selected safety analyses will be conducted by utilizing safety data from first dose through the date of the 40-week visit. Some safety analyses may be conducted after excluding data after initiation of new antihyperglycemic therapy. For rare events (<10 patients have the events), summary tables may not be generated, and individual patient level data will be listed.

Unless specified otherwise, comparisons of tirzepatide doses to placebo will be performed.

For selected continuous safety parameters, differences among treatment mean changes from baseline at scheduled visits will be assessed via an MMRM using REML. The model will include treatment, visit, treatment-by-visit interaction, country/pooled country, baseline HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]), and baseline metformin use (Yes/No) 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 covariance structures specified in Section 6.12.1.1 will be tested in order.

For selected continuous safety parameters only assessed at baseline and primary endpoint, differences among treatment mean changes from baseline in continuous safety parameters at primary endpoint will be assessed via an ANCOVA model. The model will include treatment, country/pooled country, baseline HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]), and baseline metformin use (Yes/No) as fixed effects, and baseline value of the safety parameter as a covariate.

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

Where specified, the rate of events will be analyzed using a generalized linear mixed-effects model assuming the number of events follow a negative binomial distribution and including treatment as a fixed effect. The logarithm of years in specified time interval will be adjusted as an offset to account for possible unequal treatment duration in specified time interval between patients.

6.13.1. Adverse Events

If applicable, a listing of AEs occurring after the patient's last date of study participation will be provided. The listing will include patient identification including the treatment, site number, and event information: AE group identification, event start date, Medical Dictionary for Regulatory Activities (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 the last dose of study drug to the 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 the first dose. 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 percentages will include only patients from the given sex.

A summary by treatment will be produced for the number and percentage of patients with TEAEs, serious adverse events (SAEs), death, discontinuation from study or from study treatment due to an AE will be summarized by treatment. The percentages of patients with TEAEs, overall and common (common TEAEs occurred in ≥5% of patients before rounding), 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 be shown 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:

- death
- serious adverse event,
- permanent discontinuation of study treatment due to an AE, or
- severe adverse events of special interest

6.13.1.1. Deaths

A listing of deaths will be provided. The listing will include patient identification including the treatment, site number, date of death, age at the time of enrollment, sex, MedDRA PT of associated 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 the investigator, and cause of death as adjudicated by the clinical endpoint committee (CEC).

6.13.1.2. Other Serious Adverse Events

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

A listing of all SAEs will be provided. The listing will include treatment, patient identification including the site number, treatment group, date of the event, age at the time of enrollment, sex, MedDRA SOC and PT, severity, action taken, outcome, relationship to study drug, time from first dose of study drug to the event and event duration.

6.13.1.3. Discontinuation from the Study Due to an Adverse Event

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

6.13.1.4. Discontinuation from Study Drug due to an Adverse Event

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

6.13.1.5. Treatment of Overdose

A listing of patients reporting over-dosing of tirzepatide will be provided as a protocol deviation.

6.13.2. Special Safety Topics

6.13.2.1. Hypoglycemic Events

Definitions of different categories of hypoglycemic events are included in Table GPGI.6.3.

Table GPGI.6.3. Definitions of Hypoglycemic Event Categories

	Symptoms and/or Signs of Hypoglycemia	Blood Glucose Level
Glucose Alert Value (Level 1):		
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 (Level 2):		

Definitions of Hypoglycemic Event Categories

	Symptoms and/or Signs of Hypoglycemia	Blood Glucose Level
Documented symptomatic hypoglycemia	Yes	<54 mg/dL (3.0 mmol/L)
Documented asymptomatic hypoglycemia	No	<54 mg/dL (3.0 mmol/L)
Documented unspecified hypoglycemia	Unknown	<54 mg/dL (3.0 mmol/L)
Severe Hypoglycemia (Level 3)		

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 an SAE.

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, defined in Table GPGI.6.3, occurring within a 1-hour period may be considered to be a single hypoglycemic event. Severe hypoglycemia will be considered as AESIs.

A listing of level 2 and level 3 hypoglycemic events will be provided. Statistical summaries and analyses will exclude hypoglycemic events occurring after initiation of a new antihyperglycemic therapy. For severe hypoglycemia and level 2 hypoglycemia/severe hypoglycemia incidence as well as rate per year of exposure will be provided by treatment at specified time intervals. If data warrants, additional statistical analyses will be conducted on hypoglycemic incidence and rate.

The incidence of hypoglycemic event will be analyzed using logistic regression with treatment, baseline HbA1c category, country/pooled country, and baseline metformin use (Yes/No) as fixed effects at specified time intervals. The rate of hypoglycemic episodes per patient year may be analyzed using a generalized linear mixed-effects model assuming the number of hypoglycemic episodes follows a negative binomial distribution with the mean modeled using country/pooled country, baseline metformin use (Yes/No), baseline HbA1c category, and treatment as fixed effects at specified time intervals. When the number of hypoglycemic events is less than 10, a listing of hypoglycemic events will be provided instead. The logarithm of days in specified time interval will be adjusted as an offset to account for possible unequal duration in specified time interval between patients.

6.13.2.2. Severe Persistent Hyperglycemia

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

6.13.2.3. 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 pre-defined SMQ search for acute pancreatitis and MedDRA PT of pancreatitis chronic. Detailed

searching criteria can be found in Appendix 3. Treatment-emergent adjudication-confirmed pancreatitis will be considered as AESI.

6.13.2.3.1. Pancreatic Enzyme Assessment

Observed pancreatic enzyme data (p-amylase and lipase) will be summarized by treatment and nominal visit. The number and proportion of patients with maximum postbaseline pancreatic enzyme values 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 (postbaseline measure/baseline measure) response variable and country/pooled country, baseline metformin use (Yes/No), baseline HbA1c category (≤8.0%, >8.0% [≤64, > 64 mmol/mol]), treatment, visit, and treatment-by-nominal visit interaction as fixed effects, and baseline value as a covariate.

6.13.2.4. Thyroid Malignancies, and C-Cell Hyperplasia

Treatment-emergent thyroid malignancies and C-cell hyperplasia will be identified using pre-defined MedDRA High Level Terms (HLTs) of thyroid neoplasms, and PT of thyroid C-cell hyperplasia. Detailed searching criteria can be found in Appendix 3. A summary by treatment and PT/PT within HLT will be provided. Thyroid malignancies and C-cell hyperplasia will be considered as AESI.

6.13.2.4.1. Calcitonin

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

6.13.2.5. 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 3. A summary by treatment and PT within SMQ and a listing of TEAEs will be provided. Malignancy will be considered as an AESI.

6.13.2.6. Major Adverse Cardiovascular Events

Major adverse cardiovascular events (MACE) reported by investigators are adjudicated by an independent CEC in a blinded fashion. The MACE of special interest include: 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 adjudicated as deaths due to undetermined cause by the CEC will be considered as deaths due to cardiovascular cause in statistical analysis.

A listing of patients reporting MACE, either reported by investigator or identified by the CEC, will be provided. The 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, and time from the last dose to the event (if patient has discontinued study drug prior to the event). Only adjudication-confirmed MACE will be considered as AESI.

6.13.2.7. Supraventricular Arrhythmias and Cardiac Conduction Disorders

The AE database will be searched using pre-defined SMQ or MedDRA HLT to identify events consistent with supraventricular arrhythmias and cardiac conduction disorders. Detailed searching criteria can be found in Appendix 3. Incidence of the 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 as AESIs.

6.13.2.8. Hypersensitivity Events

Two main analyses are performed for hypersensitivity reactions and related information

- **Potential Immediate Hypersensitivity:** Analysis of 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, the events occurred on the same date as the study drug injection date will be included.
- **Potential Non-Immediate Hypersensitivity:** Analysis of TEAEs occurring more than 24 hours after the end of study drug administration, but prior to subsequent study drug administration.

All potential hypersensitivity reactions will be reported by PT within SMQ decreasing frequency by treatment. The AE database will be searched using pre-defined SMQs to identify events consistent with hypersensitivity events. Detailed searching criteria for hypersensitivity events can be found in Appendix 3. Severe/serious hypersensitivity identified by pre-defined SMQ search will be considered as AESIs.

6.13.2.9. Injection Site Reactions

Injection site reactions, incidence, and related information reported via the "Injection Site Reactions" eCRF will be summarized by treatment. Information to be summarized includes the timing of the 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 HLTs of injection site reactions, administration site reactions, and infusion related reactions. Detailed searching criteria for injection site reaction events can be found in Appendix 3.The PT will be used for the summary within each HLT category. Only the severe/serious injection site reactions will be considered as AESI.

6.13.2.10. 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 (1:10) 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, with neutralizing TE ADA, and with cross-reactive TE ADA to tirzepatide will be tabulated by dose, where proportions are relative to the number of patients who are TE ADA evaluable. The frequency and percentage of patients with hypersensitivity and injection site reaction TEAEs by TE ADA status will be tabulated if data warrant.

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.

A listing may be provided for all participants who had ADA present at any time (including at baseline) or had any hypersensitivity or injection site reaction TEAE.

6.13.2.11. 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 and a summary of change from baseline in dilated fundoscopic exam will be summarized by treatment and PT.

The cases with repeat fundoscopy during the course of the trial, based on clinical suspicion of worsening retinopathy that have either findings of de novo retinopathy or progression of retinopathy, and severe/serious adverse events from the PTs defined in searching criteria in Appendix 3 will be considered as AESI and summarized.

6.13.2.12. Hepatic Safety

6.13.2.12.1. Hepatobiliary Disorders

The AE database will be searched using SMQs to identify events consistent with hepatobiliary disorders. Detailed searching criteria can be found in Appendix 2. **A summary by treatment and PT within SMQ will be provided. Severe/serious hepatobiliary disorders will be considered as AESI.**

6.13.2.12.2. 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 3. Severe/serious acute gallbladder disease will be considered as AESIs.

6.13.2.12.3. Liver Enzymes

Analyses for laboratory analyte measurements are described in Section 6.13.5. This section describes additional analyses of liver enzymes. The following will be provided by treatment group:

- A shift table of maximum to maximum alanine aminotransferase (ALT) measurement from baseline (≤1 × ULN, > 1 × ULN) to postbaseline with the following categories: ≤1×ULN, >1 to <3 × ULN, ≥3 to <5 × ULN, ≥5 to <10 × ULN, ≥10 ×ULN.
- A shift table of maximum to maximum aspartate transaminase (AST) measurement from baseline (≤1 × ULN, > 1 × ULN) to postbaseline 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 postbaseline with the following categories: ≤1×ULN, >1 to <2 × ULN, ≥2 × ULN.
- Shift tables of serum alkaline phosphatase (ALP) from baseline to postbaseline with the following categories: ≤1×ULN, >1 to <2 × ULN, ≥2 × ULN.

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

6.13.2.13. 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. PTs with severe/serious cases in the gastrointestinal SOC will be considered as AESIs.

6.13.2.14. Acute Renal Events

Laboratory measures related to renal safety will be analyzed as specified for laboratory measurements in Section 6.13.5.

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

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 3. Severe/serious acute renal events will be considered as AESIs.

6.13.2.15. Dehydration

The AE database will be searched using SMQ of dehydration to identify events consistent with dehydration. Detailed searching criteria can be found in Appendix 3. Severe/serious dehydration events will be considered as AESIs.

6.13.2.16. Metabolic Acidosis, Including Diabetic Ketoacidosis

The AE database will be searched using MedDRA PT to identify events consistent with metabolic acidosis, including diabetic ketoacidosis. Detailed searching criteria can be found in Appendix 3. The incidence of the resulting TEAEs will be summarized by treatment and PT. Severe/serious metabolic acidosis, including diabetic ketoacidosis will be considered as AESIs.

6.13.2.17. Amputation/Peripheral Revascularization

The AE database will be searched using MedDRA PT to identify events consistent with amputation or peripheral revascularization. The incidence of the resulting TEAEs will be summarized by treatment and PT. Amputation/Peripheral Revascularization will be considered as AESIs.

6.13.2.18. Major Depressive Disorder/Suicidal Ideation

The AE database will be searched using SMQs to identify events consistent with major depressive disorder or suicidal ideation. Detailed searching criteria can be found in Appendix 3. The incidence of the resulting TEAEs will be summarized by treatment and PT. Severe/serious major depressive disorder/suicidal ideation or behavior will be considered as AESIs.

6.13.3. Vital Signs

Descriptive summaries by treatment and by nominal visit will be provided for baseline and postbaseline values as well as change from baseline values. If 2 records are taken at the same visit, they will be averaged prior to being used for data summaries and analyses.

An MMRM using REML will be used to fit the changes from baseline in vital signs at all scheduled postbaseline visits. The model will include treatment, visit, treatment-by-visit interaction, country/pooled country, baseline HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]), and baseline metformin use (Yes/No) as fixed effects, and baseline value of the dependent variable as a covariate.

Counts and percentages of patients with abnormal sitting systolic BP, sitting diastolic BP, and pulse will be presented by treatment. The criteria for identifying patients with treatment-emergent vital sign abnormalities are stated in Table GPGI.6.4.

Table GPGI.6.4. Categorical Criteria for Abnormal Blood Pressure and Pulse Measurements

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

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

6.13.4. Electrocardiograms

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

The criteria for identifying patients with treatment-emergent quantitative ECG abnormalities is outlined in Table GPGI.6.5.

In addition, the percentage of patients with QT greater than 500 msec will be summarized, and the percentage of patients with QTcF greater than 500 msec will also be summarized (refer to PSAP).

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

Table GPGI.6.5. 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

Selected Categorical Limits for ECG Data
Abbreviations: bpm = beats per minute; ECG = electrocardiogram; QTCF = Fridericia's corrected QT interval.

6.13.5. Clinical Laboratory Evaluation

All laboratory data will be reported in the International System of Units and Conventional Units. Values that are outside of reference ranges will be flagged as high (H) or low (L) in the listings. 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 measurements.

Observed 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.

Shift tables will be produced for selected measurements. A shift table will include unplanned 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. The proportion of patients shifted will be compared between treatments.

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

6.14. Health Outcomes

The patient-reported outcome questionnaires will be completed by the patients at baseline and at 40 weeks (or early termination visit prior to 40 weeks). Main analyses will be performed on the EAS and sensitivity analyses on the FAS. Using ANCOVA analyses, p-values and 2-sided 95% CIs of health outcome measures described below will be derived and summarized for the 5 mg, 10 mg, and 15 mg doses of tirzepatide compared to placebo.

No multiplicity adjustment will be made in the evaluation of health outcome measures. Item-level missingness is dealt with as per the instrument developer's instruction.

6.14.1. EQ-5D-5L

A descriptive frequency table of individual items in EQ-5D-5L questionnaire will present baseline, observed endpoint, and endpoint including last observation carried forward (LOCF, exclude baseline observation) values as separate summaries. The changes from baseline to Week 40, with and without LOCF of 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 (≤8.0%, >8.0% [≤64, >64 mmol/mol]) and baseline metformin use (Yes/No) as fixed effects, and corresponding baseline EQ-5D-5L score as a covariate.

6.14.2. Impact of Weight on Self-Perceptions Questionnaire

A descriptive frequency table of individual items in IW-SP questionnaire will be presented at baseline, observed endpoint, and endpoint including last observation carried forward (LOCF, exclude baseline observation). The changes from baseline to Week 40, with and without LOCF, of the raw and transformed total IW-SP scores will be analyzed using an ANCOVA model with model terms of treatment, country/pooled country, baseline HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]) and baseline metformin use (Yes/No) as fixed effects, and corresponding baseline IW-SP score as a covariate.

6.14.3. Ability to Perform Physical Activities of Daily Living

A descriptive frequency table of individual items in APPADL questionnaire will be presented at baseline, observed endpoint, and endpoint including last observation carried forward (LOCF, exclude baseline observation). The changes from baseline to Week 40, with and without LOCF, of the of the raw and transformed total APPADL scores will be analyzed using an ANCOVA model with model terms of treatment, country/pooled country, baseline HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]) and baseline metformin use (Yes/No) as fixed effects, and corresponding baseline APPADL score as a covariate.

6.14.4. Diabetes Treatment Satisfaction Questionnaire

Descriptive summaries will be provided at baseline (DTSQs only),observed endpoint, and endpoint including last observation carried forward excluding baseline observations (DTSQc only) for the perceived hyperglycemia item, perceived hypoglycemia item, and 6-item overall satisfaction score.

The DTSQc score at Week 40, with and without LOCF, will be analyzed using an ANCOVA model with model terms of treatment, country/pooled country, baseline HbA1c (≤8.0%, >8.0% [≤64, >64 mmol/mol]), and baseline metformin use (Yes/No) as fixed effects, and baseline DTSQs score as a covariate. The analyses will be conducted for the perceived hyperglycemia item, perceived hypoglycemia item, and 6-item overall satisfaction score.

6.15. Subgroup Analyses

Efficacy subgroup analyses will be guided by the efficacy estimand. Additional subgroup analyses guided by the treatment-regimen estimand will be performed for HbA1c.

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.

6.15.1. Subgroup Analysis of HbA1c Change at 40 Weeks

Subgroup analyses by the following baseline characteristics will be provided: age group (<65, ≥65 years), age group (<75, ≥75 years), race, gender, ethnicity, baseline metformin use (Yes, No), region of enrollment (US, Japan, EU), duration of diabetes (<median, ≥median), duration of diabetes (≤5, >5 to ≤10, >10 years), HbA1c (≤8.0%, >8.0%), renal impairment (eGFR <60, ≥60 mL/min/1.73m^2), BMI group (<27, ≥27 kg/m^2), and BMI group (<30, ≥30 to <35, ≥35 kg/m^2).

6.15.2. Subgroup Analysis of Weight Change at 40 Weeks

Subgroup analyses by the following baseline characteristics will be provided: age group (<65, ≥65 years), age group (<75, ≥75 years), race, gender, region of enrollment (US, Japan, EU), duration of diabetes (<median, ≥median), HbA1c (≤8.0%, >8.0%), renal impairment (eGFR <60, ≥60mL/min/1.73m^2), BMI group (≤27, >27), and BMI group (<30, ≥30 to <35, ≥35).

6.16. Interim Analyses and Data Monitoring Committee

No interim analyses are planned for this study.

6.17. 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.

6.17.1. Patients Impacted by COVID-19

Listings of patients with protocol deviation or mitigation due to COVID-19/ related restrictions, patients with COVID-19 adverse events or death, and patient's dispositions with reasons related to COVID-19 will be provided. Listing will include patient identification, treatment, date of impact, and description of impact.

6.17.2. Adverse Events

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

6.17.3. Patient Disposition

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

6.17.4. Study Visits

A summary of patients with study visit impacted by COVID-19/ related restrictions 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.

6.17.5. 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. Additional summary may be provided by country of enrollment and treatment group.

6.17.6. 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.

7. Unblinding Plan

The unblinding plan will be located in a separate document.

8. References

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

9. Appendices

Appendix 1. SAS Code for Multiple Imputation

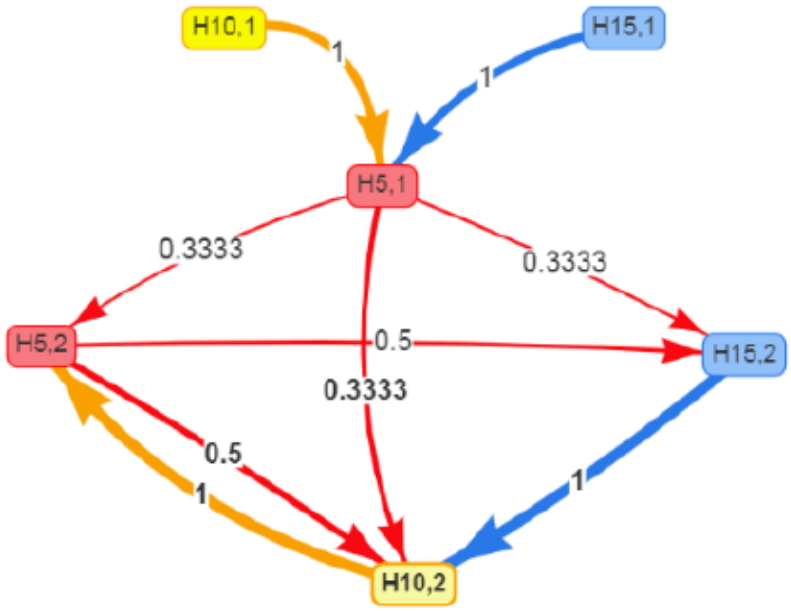
```
*retrieved dropout imputation*;
*ret fl is a flag variable for subjects that are retrieved dropouts*;
proc mi data=&dataset_in. seed=&seed. out=&dataset_out. nimpute=&n_impute.;
by trtpn;
class ret_fl;
monotone reg(chg/details);
mnar model (chg/modelobs=(ret_fl='Y'));
var base chg;
run;

*placebo imputation*;
*replace 'placebo' with the assigned value for the placebo group*;
proc mi data=&dataset_in. seed=&seed. out=&dataset_out_2. nimpute=&n_impute.;
class trtpn;
monotone reg(chg/details);
mnar model (chg/modelobs=(trtpn='placebo'));
var base chg;
run;
```

Appendix 2. Statistical Analysis for Japan

Statistical analysis methods for the primary efficacy analysis for the Pharmaceuticals and Medical Devices Agency (PMDA) is same as described in Section 6.12.1. For the PMDA, the primary efficacy assessment will be guided by the "efficacy" estimand, with controlling overall family-wise type 1 error rate at a 2-sided alpha of 0.05 only for primary endpoint and body weight evaluation that tirzepatide once weekly (QW) is superior to placebo in hemoglobin A1c (HbA1c) and body weight change from baseline to 40 weeks. The primary objective hypotheses and the key secondary for body weight for the PMDA are as follows:

1. $H_{5,1}$, $H_{10,1}$, and $H_{15,1}$: Superiority test of tirzepatide 5 mg, 10 mg, 15 mg versus placebo in HbA1C change from baseline at 40 weeks respectively.
2. $H_{5,2}$, $H_{10,2}$, and $H_{15,2}$: Superiority test of tirzepatide 5 mg, 10 mg, 15 mg versus placebo in body weight change from baseline at 40 weeks respectively.

The statistical testing scheme for the PMDA is described in the following diagram. $H_{10,1}$ and $H_{15,1}$ will be initially tested each at 0.025 significance level.



Subgroup analyses for patient disposition, patient characteristics, key efficacy and safety measures by country (Japanese/non-Japanese) will be provided to support Japan regulatory registration.

Appendix 3. Searching Criteria for Adverse Events of Special Interest

The adverse event of special interest (AESI) analyses are detailed in Section 6.13.2. The search criteria for each AESI are stored in CLUWE:

CCI [redacted]

Leo Document ID = 59721d0f-8876-48d4-a33b-746bac6ecc8d

Approver: PPD

Approval Date & Time: 01-Feb-2021 17:59:43 GMT

Signature meaning: Approved