Statistical Analysis Plan: I8F-MC-GPGV

A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study with an Open-Label Extension Assessing the Efficacy, Safety, and Pharmacokinetics/Pharmacodynamics of Tirzepatide in Pediatric and Adolescent Participants with Type 2 Diabetes Mellitus Inadequately Controlled with Metformin, or Basal Insulin, or Both (SURPASS-PEDS)

NCT05260021

Approval Date: 15-Jan-2025

Title Page

Protocol Title: A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study with an Open-Label Extension Assessing the Efficacy, Safety, and Pharmacokinetics/Pharmacodynamics of Tirzepatide in Pediatric and Adolescent Participants with Type 2 Diabetes Mellitus Inadequately Controlled with Metformin, or Basal Insulin, or Both (SURPASS-PEDS)

Protocol Number: I8F-MC-GPGV

Compound Number: LY3298176

Short Title: A Study to Evaluate Tirzepatide Compared to Placebo in Pediatric and Adolescent Participants with Type 2 Diabetes Mellitus Inadequately Controlled with Metformin, or Basal Insulin, or Both

Acronym: SURPASS-PEDS

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana, USA 46285
Regulatory Agency Identifier Number(s) IND: 128801

Document ID: VV-CLIN-140059

Approved on 15 Jan 2025 GMT

Table of Contents

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Version history

SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1	12 Jan 2023	Not Applicable	Original version
2	25 Sep 2024	Section 1.1*: Added new endpoint percentage change in BMI, to key secondary endpoint	Additional key secondary endpoint was added to assess clinically relevant changes in BMI.
		Section 1.1*: Added new endpoints as additional secondary objectives: • Incidence of BMI reduction of ≥5%, ≥10%, ≥ 15% • Percent change from baseline and change from baseline in body weight • Height Velocity	Additional secondary endpoints were added to assess clinically relevant changes in weight, BMI and height velocity.
		Section 1.1*: Updated the exploratory objectives as: • Incidence of HbA1c ≤6.5% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia • Incidence of HbA1c < 7.0% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia	The exploratory objectives were updated to reflect assessment of achievement of a clinically relevant HbA1c without clinically significant hypoglycemia.
		Section 1.1*: Updated and clarified the text of estimands	Language was modified for Clarification.
		Section 1.2: Updated and clarified the text for study design	Language was modified for clarification.
		Section 2*: Added hypotheses for newly added key secondary endpoint and updated graphical testing scheme	Graphical testing strategy was revised per additional key secondary endpoints subject to type 1 error rate control.

		Section 3*: • Updated the mITT population include inadvertently enrolled population in analysis • Added tirzepatide-treated analyses definition in SS2	Analysis sets definitions were updated for clarity.
		Section 4: • Updated text to mention at Section 4.1.4 – "There is no comparison between any of the treatment group at week 52." • Updated Section 4.1.5 specifying model for week 30 and week 52 as well for analytical approach efficacy estimand • Updated analytical approach for treatment regimen estimand, "Primary Multiple imputation method as J2R with multiple imputations."* • Updated analytical approach for binary endpoints – using logistic regression and presenting relative risk and risk difference instead of "longitudinal logistic regression" and imputing missing data assuming "Missing at Random."* • Updated Analysis Related to safety estimand by providing models and detailed analyses approach • Update the missing data imputation – will refer to J2R instead of retrieved dropouts as "Primary Multiple imputation" method and multiple imputation method for efficacy estimand* • Provided baseline definition for concomitant therapy and update the definition for rescue therapy • Section 4.6 - Text corrections, provided definition to tirzepatide treated analysis, SMQ numbers LLT codes, PT's, elaborated detailed analysis methods on Immunogenicity. And updated the analyses methods and analyses sets for safety analyses. • Section 4.7 – updated the analysis methods of Health outcomes from ANCOVA to MMRM analysis, provided references, textual corrections; modified the subgroup analyses	Details of analysis methods were updated per the changes of objectives, corrected for clarity and accuracy, and aligned with the standardized analysis approach across tirzepatide indications.
		Section 6: • Added search criteria's for AESIs • Added a subsection about Important Protocol Deviations • Provided table for parameters need to be included in Baseline Demographics • Added a subsection on clinical trial registry	Updated to align with the standard analysis approach across tirzepatide indications.
		Section 7: • Added reference	Updated references as per the changes in respective sections.

3	See date on page 1	Section 2: graphical testing scheme was revised.	Graphical testing strategy was revised subject to type 1 error rate control.
		Section 4.1.2: updated the text regarding which estimand is used for newly added additional secondary endpoints at Week 30 and Week 52 objectives. Section 4.1.5.2: updated the text for analyses at Week 52. Section 4.1.5.3 and 4.1.5.4: removed model statement at Week 52, since there is no treatment comparison at Week 52. Section 4.1.6: updated the missing data imputation method for analyses at Week 52 based on treatment regimen estimand. Section 4.5: added treatment regimen estimand for additional secondary analyses at Week 30 and at Week 52.	Updated the analysis approach as per design and addressed FDA comments and PIP requirements. Updated to align with the standard analysis approach across tirzepatide indications
		Section 4.6.4: updated the text and clarified the references for standard deviation score (SDS). Section 4.6.5 and 4.7.1: updated and clarified the text of analyses periods. Section 4.6.6.12: updated definition of albuminuria. Section 4.7.2.2*: updated the questionnaire version and dimensions.	Language was modified for clarification.
		Section 4.7.3: added safety subgroup analyses.	Updated to align with the standard analysis approach across tirzepatide indications.
		Section 5*: text was updated by removing the definition of evaluable participants.	To align with pediatric investigation plan (PIP)
		Section 6: updated SMQ's and PT's as per the latest MedDRA version and added few parameters added to appendix 3.	Minor textual correction and updates

* Indicates the changes to protocol-planned analyses.

Abbreviations and Definitions

Term	Definition
ACR	albumin/creatinine ratio
ADA	anti-drug antibodies
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BG	blood glucose
BMI	body mass index
BP	blood pressure
CDC	Centers for Disease Control and Prevention
CI	confidence interval
CN	conventional units of measurement
CWM	chronic weight management
DBL	database lock
DBP	diastolic blood pressure
DC	Discontinuation
DMC	Data Monitoring Committee
EAS	efficacy analysis set
ECG	Electrocardiogram
eCRF	electronic case report form
eGFR	estimated glomerular filtration rate
FAS	full analysis set
FSG	fasting serum glucose

Term	Definition
GI	Gastrointestinal
HbA1c	glycated hemoglobin A1C
HDL	high-density lipoprotein
HLT	High Level Term
ID	Identification
J2R	jump to reference
LDL	low-density lipoprotein
LLT	Lowest Level Term
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MH	medical history
mITT	modified intent-to-treat
MMRM	mixed model for repeated measures
MNAR	missing not at random
MRD	minimum required dilution
NAb	neutralizing antibodies
PD	pharmacodynamic(s)
PedsQL	Pediatric Quality of Life
PIP	Pediatric Investigation Plan
PK	pharmacokinetic(s)
PRO	patient reported outcome
PT	Preferred Term
QTc	corrected QT interval
QTcF	Fridericia's corrected QT interval
QW	once weekly
SAE	serious adverse event

Term	Definition
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	standard deviation
SDS	standard deviation score
SI	Système International units
SMQ	Standardized MedDRA Query
SOC	System Organ Class
SS1	safety analysis set 1
SS2	safety analysis set 2
T2DM	type 2 diabetes mellitus
TE	treatment-emergent
TE ADA	treatment-emergent anti-drug antibodies
TEAE	treatment emergent adverse event
TG	Triglyceride
TZP	Tirzepatide
UACR	urine albumin-to-creatinine ratio
ULN	upper limit of normal
VAS	Visual Analog Scale
VLDL	very low-density lipoprotein
WHO	World Health Organization

1. Introduction

This document is the SAP for Study I8F-MC-GPGV (SURPASS-PEDS). SURPASS-PEDS is a randomized, double-blind, parallel arm, placebo-controlled study with an open-label extension, to assess the efficacy, safety, and PK/PD of tirzepatide in male and female participants 10 to <18 years of age with T2DM and inadequate glycemic control on diet and exercise with metformin, or basal insulin, or both.

Tirzepatide is a glucose-dependent insulinotropic polypeptide and a glucagon-like peptide-1 receptor agonist that is administered once weekly. Tirzepatide is approved for T2DM and CWM in adults and is being investigated for its potential use in the treatment of T2DM and CWM in the pediatric and adolescent population.

T2DM is a disease that is primarily diagnosed in adults, and the risk of the disease increases with age (CDC 2020). There has been a significant increase of T2DM in the pediatric and adolescent population in recent years, although the absolute number of youths with T2DM remains low and type 1 diabetes mellitus accounts for approximately 90% of all diabetes in the pediatric population (Dabelea et al. 2014).

The purpose of this study is to evaluate the safety and efficacy of tirzepatide 5 mg and tirzepatide 10 mg compared to placebo, with metformin, or basal insulin, or both, for glycemic control in the pediatric and adolescent population with T2DM over a 30-week period, followed by a 22-week open-label extension. The starting dose of tirzepatide is 2.5 mg followed by a dose escalation of 2.5-mg increments every 4 weeks to achieve the assigned maintenance dose. The types of data will include efficacy, safety, immunogenicity, PK, and PD data. At least 90 participants will be randomized in a 1:1:1 ratio to tirzepatide 5 mg, tirzepatide 10 mg, or placebo.

This SAP supersedes all statistical considerations and analyses described in Protocol GPGV.

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
• To demonstrate that pooled 5 and 10 mg tirzepatide QW is superior to placebo at 30 weeks	• Change in HbA1c from baseline
Key Secondary (Controlled for Type I Error)	
• To demonstrate that 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, is superior to placebo at 30 weeks	• Incidence of HbA1c ≤6.5% • Change in BMI SDS (age- and sex-matched) from baseline • Change in FSG from baseline • Percent change in BMI from baseline
• To demonstrate that 5 and 10 mg tirzepatide QW, analyzed as separate treatment arms, is superior to placebo at 30 weeks	• Change in HbA1c from baseline
Additional Secondary (Not Controlled for Type I Error)	
• To demonstrate that 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, is superior to placebo at 30 weeks	• Incidence of HbA1c <7.0% • Incidence of HbA1c <5.7% • Change in daily average 6-point self-monitored blood glucose profiles from baseline
• To assess the effect of 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, at 52 weeks	• Change in HbA1c from baseline • Incidence of HbA1c <7.0% • Incidence of HbA1c ≤6.5% • Incidence of HbA1c <5.7% • Change in FSG from baseline
• To demonstrate that 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, is superior to placebo at 30 weeks	• Change in BMI from baseline • Incidence of BMI reduction ≥5% • Incidence of BMI reduction ≥10% • Incidence of BMI reduction ≥15% • Percentage change in body weight from baseline • Change in body weight from baseline • Change in waist circumference from baseline
• To assess the effect of 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, at 52 weeks	• Change in BMI from baseline • Percent change in BMI from Baseline • Change in BMI SDS (age- and sex-matched) from baseline • Incidence of BMI reduction ≥5% • Incidence of BMI reduction ≥10% • Incidence of BMI reduction ≥15% • Percentage change in body weight from baseline • Change in body weight from baseline • Change in waist circumference from baseline

Objectives	Endpoints
• To compare 5 mg, and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, to placebo at 30 weeks, and • To assess the effect of 5 mg and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms at 52 weeks	• Incidence of new or worsening of dyslipidemia • Incidence of new or worsening of hypertension • Incidence of new or worsening of albuminuria • Incidence of new or worsening of diabetic neuropathy • Incidence of new or worsening of diabetic retinopathy or diabetic maculopathy in either eye
• To assess safety of 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, at 30 weeks and through end of the safety follow-up period	• Treatment-emergent adverse events • Discontinuation of study intervention due to AEs • Adjudicated pancreatic AEs • Serum calcitonin • Incidence of allergic, hypersensitivity reactions, and injection site reactions • Incidence of treatment-emergent anti-drug antibodies to tirzepatide • Percentage change from baseline for serum lipid levels • 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
• To compare 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, to placebo at 30 weeks, and • To assess the effect of 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, at 52 weeks or end of the safety follow-up period	• Change from baseline for physical and developmental measures o height o height SDS and weight SDS o height velocity o Tanner Staging
• Characterize the pharmacokinetics and pharmacodynamic relationships of tirzepatide	• Relationship between tirzepatide exposure and key safety and efficacy measures
• To assess the effect of 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, on patient-reported outcomes at 30 and 52 weeks	• Change from baseline for o PedsQL Generic Core Scales o PedsQL (3.2) Diabetic Module from baseline

Objectives	Endpoints
Tertiary/Exploratory	
• To compare 5 mg and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, to placebo at 30 weeks, and • To assess the effect of 5 mg and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, at 52 weeks	• Incidence of HbA1c ≤6.5% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia • Incidence of HbA1c <7.0% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia
• To assess the effect of 5 and 10 mg tirzepatide QW, analyzed as pooled and separate treatment arms, on patient-reported outcomes at 30 and 52 weeks	• Change from baseline for EQ-5D-Y
• To assess the effect of 5 and 10 mg tirzepatide QW, separately and pooled, at Weeks 8, 16, and 30 on additional metabolic measures	• Measures of insulin resistance, alpha cell and beta cell function, and serum adiponectin

Abbreviations: AEs = adverse events; BMI = body mass index; FSG = fasting serum glucose; HbA1c = glycated hemoglobin ; PedsQL = Pediatric Quality of Life Inventory; QW = once weekly; SDS = standard deviation score.

Primary estimands

The primary and each key secondary efficacy analysis will be guided by the "efficacy" estimand and the "treatment regimen" estimand.

Efficacy estimand

The clinical question of interest for the efficacy estimand is the treatment difference between tirzepatide and placebo after 30 weeks of intervention in pediatric and adolescent participants with T2DM, prior to intervention discontinuation for any reason or initiation of rescue antihyperglycemic medication.

Efficacy estimand attributes

- *Population:* Pediatric and adolescent participants with study condition/disease who received at least 1 dose of study treatment.
- *Treatment condition:* The randomly assigned treatment without the influence of rescue antihyperglycemic medication.
- *Endpoints:* The primary and key secondary endpoints will be studied. Further details on the endpoints can be found in the Objectives and Endpoints table (Section 1.1).
- *Population level summary:* The difference in mean change from baseline to 30 weeks will be used for continuous endpoints; the difference in proportion (absolute or relative, as appropriate) will be used for dichotomous endpoints. The population level summary will be conducted using the EAS1.

- *Intercurrent events:* The 2 intercurrent events "intervention discontinuation for any reason" and "initiation of rescue medication" are both addressed by the treatment condition of interest attribute. There are no other intercurrent events.
- *Rationale:* The efficacy estimand aims to evaluate the efficacy under the ideal condition that all participants would adhere to the randomly assigned study intervention and without being confounded by the initiation of new antihyperglycemic medications.

Treatment-regimen estimand

The clinical question of interest for the treatment-regimen estimand is the treatment difference between tirzepatide and placebo after 30 weeks of intervention in pediatric and adolescent participants with T2DM, regardless of intervention discontinuation for any reason or initiation of rescue antihyperglycemic intervention.

Treatment-regimen estimand attributes

- *Population:* Pediatric and adolescent participants with study condition/disease who received at least 1 dose of study treatment.
- *Treatment condition:* The randomly assigned treatment regardless of treatment discontinuation or influence of rescue antihyperglycemic medication.
- *Endpoints:* The primary and key secondary endpoints will be studied. Further details on the endpoints can be found in the Objectives and Endpoints table (Section 1.1).
- *Population level summary:* The difference in mean change from baseline to 30 weeks will be used for continuous endpoints and the difference in proportion (absolute or relative, as appropriate) will be used for dichotomous endpoints. The population level summary will be conducted using the FAS.
- *Intercurrent events:* There are no intercurrent events.

Rationale: The treatment-regimen estimand estimates treatment effect, including the effect of treatment discontinuation and influence of rescue medication to reflect the real-life behavior of the target population.

Safety estimand

The clinical interest for safety estimands is the safety assessment of individual treatment arms up to Week 30 and up to end of study in pediatric and adolescent participants with T2DM, from all randomly assigned participants who are exposed to at least 1 dose of study intervention, regardless of adherence to study intervention or initiation of rescue antihyperglycemic medication.

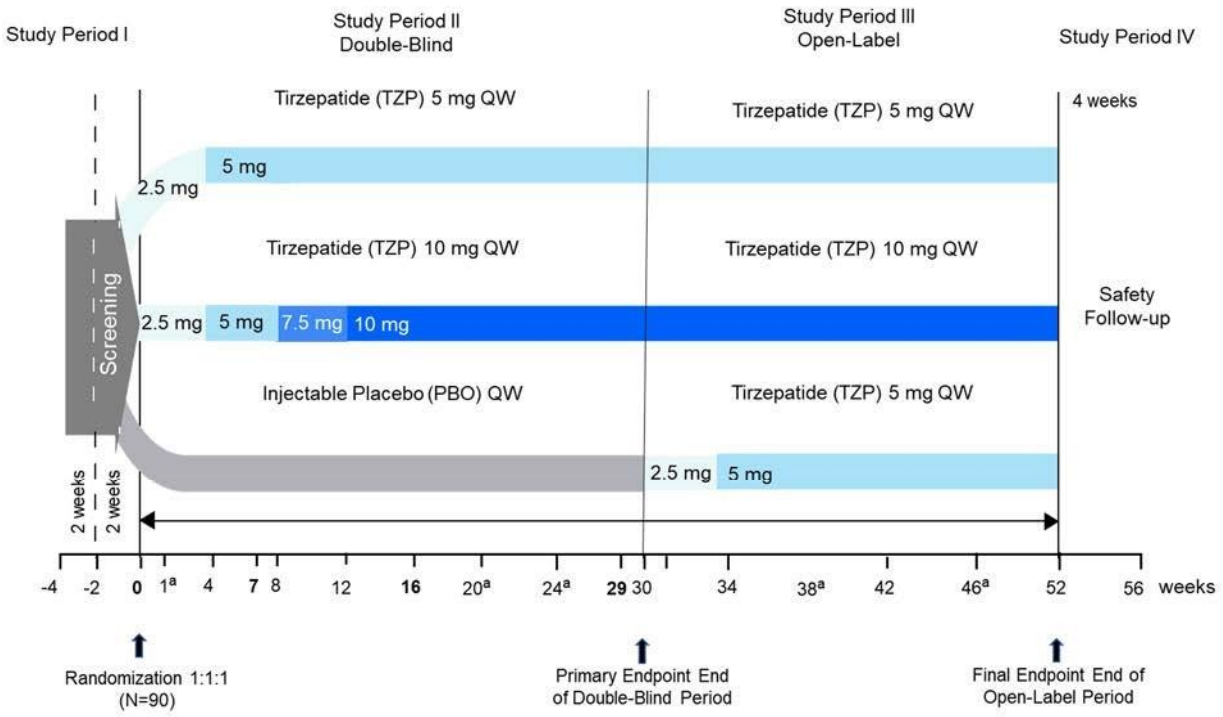
Safety estimand attributes

- *Population:* Pediatric and adolescent participants with study condition/disease who received at least 1 dose of study treatment.

- *Treatment condition:* The randomly assigned treatment regardless of treatment discontinuation or influence of rescue antihyperglycemic medication.
- *Endpoints:* Endpoints corresponding to the safety analyses described in Section 4.5.
- *Population level summary:* Population level summaries will be conducted using the SS1 for double-blind period, and SS2 for double-blind, open-label and follow-up periods.
- *Intercurrent events:* There are no intercurrent events.

1.2. Study Design

Study GPGV will evaluate 5- and 10-mg doses of tirzepatide starting at 2.5 mg and using a dose escalation of 2.5-mg increments every 4 weeks to achieve the assigned maintenance dose. The total duration of study participation for each participant post-randomization is 30 weeks of double-blind period and 22 weeks of open-label period, adding to a total of 52 weeks with an additional 4 weeks of safety follow-up.



Abbreviations: N = number of patients in the analysis population; PBO = placebo; QW = once weekly.

[a] Phone visit.

CCI [redacted]

Figure GPGV.1.1. Schema of study design.

1.2.1. Brief Summary

The purpose of this study is to evaluate the efficacy and safety of tirzepatide 5 and 10 mg compared to placebo for glycemic control in the pediatric and adolescent population with T2DM over a 30-week period, followed by a 22-week open-label extension.

1.2.2. Number of Participants

At least 90 participants will be randomized in a 1:1:1 ratio to each study arm.

Participant stratification factors are:

- age (≤ 14 years of age, >14 years of age), and
- baseline antihyperglycemic medication use of
  - metformin only, or
  - basal insulin only, or
  - metformin and basal insulin.

1.2.3. Intervention Groups and Duration

The overall study duration is approximately 60 weeks over 4 required study periods. Table GPGV.1.1 describes the study periods.

Table GPGV.1.1. Intervention Groups and Duration

Study Period	Duration	Intervention
Screening/lead-in	Approximately 4 weeks	-----
Double-blind, placebo-controlled	30 weeks	5 mg tirzepatide 10 mg tirzepatide, or placebo
Open-label extension on active treatment period	22 weeks	5 mg tirzepatide, or 10 mg tirzepatide
Post-treatment safety follow-up	30 days	-----

In the 22-week open-label period, participants in 5- and 10-mg arms will continue with current treatment while participants randomized to placebo will initiate treatment at 2.5 mg QW and after 4 weeks, will receive 5 mg of tirzepatide for the remainder of the period.

1.2.4. Data Monitoring Committee

An external DMC supported by an independent statistical analysis center, will have the responsibility to review unblinded interim analyses results to monitor the safety of the participants in the study until the last participant completes the Week 52 visit. More details are included in the DMC charter.

2. Statistical Hypotheses Subject to Type I Error Rate Control

The hypotheses relative to the primary and key secondary endpoints are whether there is a superiority in tirzepatide 5 mg (TZP 5 mg), tirzepatide 10 mg (TZP 10 mg), and tirzepatide pooled 5 mg and 10 mg (TZP_ALL), compared to placebo at 30 weeks in changes in the measures outlined in Section 1.1. Thus, the hypotheses to be tested in relation to the primary and key secondary estimands are as follows:

- **Hypotheses on the mean change in HbA1c:**
  - $H_{1,0}$: TZP_ALL is not superior to placebo in mean change from baseline in HbA1c at 30 weeks
  - $H_{1,a}$: TZP_ALL is superior to placebo in mean change from baseline in HbA1c at 30 weeks
  - $H_{2,0}$: TZP 10 mg is not superior to placebo in mean change from baseline in HbA1c at 30 weeks
  - $H_{2,a}$: TZP 10 mg is superior to placebo in mean change from baseline in HbA1c at 30 weeks
  - $H_{3,0}$: TZP 5 mg is not superior to placebo in mean change from baseline in HbA1c at 30 weeks, and
  - $H_{3,a}$: TZP 5 mg is superior to placebo in mean change from baseline in HbA1c at 30 weeks
- **Hypotheses on the incidence of HbA1c ≤6.5%**
  - $H_{4,0}$: TZP_ALL is not superior to placebo in proportion of patients achieving an HbA1c ≤6.5% at 30 weeks
  - $H_{4,a}$: TZP_ALL is superior to placebo in proportion of patients achieving an HbA1c ≤6.5% at 30 weeks
  - $H_{5,0}$: TZP 10 mg is not superior to placebo in proportion of patients achieving an HbA1c ≤ 6.5% at 30 weeks
  - $H_{5,a}$: TZP 10 mg is superior to placebo in proportion of patients achieving an HbA1c ≤6.5% at 30 weeks
  - $H_{6,0}$: TZP 5 mg is not superior to placebo in proportion of patients achieving an HbA1c ≤ 6.5% at 30 weeks, and
  - $H_{6,a}$: TZP 5 mg is superior to placebo in proportion of patients achieving an HbA1c ≤6.5% at 30 weeks
- **Hypotheses on the mean change in FSG**
  - $H_{7,0}$: TZP_ALL is not superior to placebo in mean change from baseline in FSG at 30 weeks
  - $H_{7,a}$: TZP_ALL is superior to placebo in mean change from baseline in FSG at 30 weeks
  - $H_{8,0}$: TZP 10 mg is not superior to placebo in mean change from baseline in FSG at 30 weeks
  - $H_{8,a}$: TZP 10 mg is superior to placebo in mean change from baseline in FSG at 30 weeks

  - $H_{9,0}$: TZP 5 mg is not superior to placebo in mean change from baseline in FSG at 30 weeks, and
  - $H_{9,a}$: TZP 5 mg is superior to placebo in mean change from baseline in FSG at 30 weeks
- **Hypotheses on the mean change in BMI SDS**
  - $H_{10,0}$: TZP_ALL is not superior to placebo in mean change from baseline in BMI SDS at 30 weeks
  - $H_{10,a}$: TZP_ALL is superior to placebo in mean change from baseline in BMI SDS at 30 weeks
  - $H_{11,0}$: TZP 10 mg is not superior to placebo in mean change from baseline in BMI SDS at 30 weeks
  - $H_{11,a}$: TZP 10 mg is superior to placebo in mean change from baseline in BMI SDS at 30 weeks
  - $H_{12,0}$: TZP 5 mg is not superior to placebo in mean change from baseline in BMI SDS at 30 weeks
  - $H_{12,a}$: TZP 5 mg is superior to placebo in mean change from baseline in BMI SDS at 30 weeks
- **Hypotheses on the Percent Change in BMI**
  - $H_{13,0}$: TZP_ALL is not superior to placebo in percent change from baseline in BMI at 30 weeks
  - $H_{13,a}$: TZP_ALL is superior to placebo in percent change from baseline in BMI at 30 weeks
  - $H_{14,0}$: TZP 10 mg is not superior to placebo in percent change from baseline in BMI at 30 weeks
  - $H_{14,a}$: TZP 10 mg is superior to placebo in percent change from baseline in BMI at 30 weeks
  - $H_{15,0}$: TZP 5 mg is not superior to placebo in percent change from baseline in BMI at 30 weeks
  - $H_{15,a}$: TZP 5 mg is superior to placebo in percent change from baseline in BMI at 30 weeks

For continuous endpoints (HbA1c, BMI, and FSG), the hypothesis tests will be carried out as follows. Let the mean treatment difference be defined as μ = (mean change from baseline in endpoint of TZP minus mean change from baseline in endpoint of placebo). The superiority of TZP to placebo will be tested with the hypotheses

$$H_0: \mu \geq 0.0 \text{ against } Ha: \mu < 0.0$$

For the binary endpoint, let p_T and p_P be the proportions of participants achieving HbA1c at a specified target in TZP and placebo groups respectively. The hypotheses to be tested are:

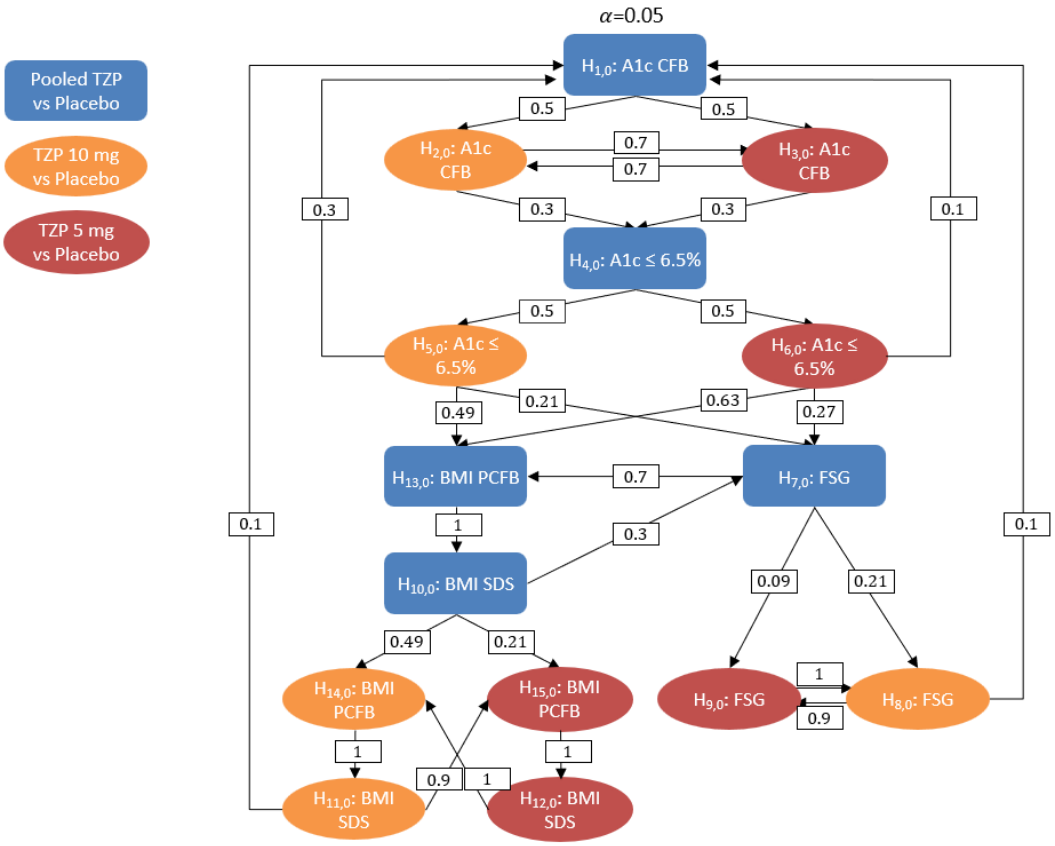
$$H_0: p_T \leq p_P \text{ against } Ha: p_T > p_P$$

2.1. Type I Error Rate Control Strategy

The 2 estimands of the primary efficacy objectives are intended for different purposes (see rationales in Section 1.1). For that reason, there is no type 1 error rate adjustments to be made for conducting analyses relative to efficacy and treatment-regimen estimands. Moreover, no multiplicity adjustments will be made for conducting separate analyses relative to additional secondary and exploratory efficacy objectives or safety assessments.

Statistical comparisons for the primary and key secondary endpoints will be carried out using a graphical approach (Bretz et al. 2009, 2011) to strongly control the type I error at a 2-sided significance level of 0.05. The graphical approach will be conducted separately for each of the estimands, and each will be tested at the full significance level of 0.05.

The type 1 error control strategy is illustrated on the graphical testing scheme on Figure GPGV.2.1.



Abbreviation: A1c = glycated hemoglobin; BMI = body mass index; CFB = change from baseline; FSG = fasting serum glucose; PCFB = percent change from baseline; SDS = standard deviation score; TZP = tirzepatide.

Figure GPGV.2.1. Graphical testing scheme.

3. Analysis Sets

For statistical analysis purposes, the following participant analysis sets are defined in Table GPGV.3.1.

Table GPGV.3.1. Participants Analysis Sets

Participant Analysis Set	Description
Screened Participants	All participants who sign the Informed Consent Form.
Randomized participants	All participants who are randomly assigned to a treatment arm.
Modified Intent-to-Treat (mITT) population	All randomly assigned participants who are exposed to at least 1 dose of study intervention (that is tirzepatide or placebo).
Efficacy Analysis Set 1 (EAS1)	Data obtained during Double-Blind Period from the mITT population, excluding data after initiating rescue antihyperglycemic medication or discontinuing from study intervention.
Efficacy Analysis Set 2 (EAS2)	Data obtained during Double-Blind Period and Open-Label Period from the mITT population, excluding data after initiating rescue antihyperglycemic medication or discontinuing from study intervention.
Full Analysis Set (FAS)	Data obtained during Double-Blind Period from the mITT population, regardless of adherence to study intervention or initiation of rescue antihyperglycemic medication.
Full Analysis Set 2 (FAS2)	Data obtained during Double-Blind Period and Open-Label Period from the mITT population, regardless of adherence to study intervention or initiation of rescue antihyperglycemic medication.
Safety Analysis Set 1 (SS1)	Data obtained during Double-Blind Period from the mITT population, regardless of adherence to study intervention or initiation of rescue antihyperglycemic medication. Note: SS1 is the same as FAS.
Safety Analysis Set 2 (SS2)	Data obtained during Double-Blind Period, Open-Label Period and Safety Follow-up Period from the mITT population, regardless of adherence to study intervention or initiation of rescue antihyperglycemic medication. For AE related analyses and lab/vitals/ECG categorical analyses, TZP-treated participants, defined as a subset of the mITT population who are exposed to at least 1 dose of tirzepatide, will be included in the analyses.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. All statistical analyses will be conducted with SAS Version 9.4 or higher unless otherwise stated.

Some analyses and summaries described in this analysis plan may not be conducted if not warranted by data (for example, few events to justify conducting an analysis). 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 final DBL.

All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05/1-sided alpha level of 0.025, unless otherwise stated, and all CIs will be given at a 2-sided 95% level. In statistical summaries and analyses, all data, including those collected during the open-label period, will be analyzed by randomized treatment assignment. In the event of a treatment error, participants will be analyzed according to the treatment they were randomly assigned to.

4.1.1. Baseline Definition

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.

- Immunogenicity data collected up to the first dose time will serve as baseline.
- Vitals, labs and ECG baseline needs to be prior to or within 1 hour after the first dose time.
- Patient-reported outcomes data collected at Visit 3, regardless of the timing relative to first dose, will serve as baseline.

4.1.2. Analysis Sets and Estimands

The primary efficacy assessment, guided by the *efficacy estimand*, will be conducted using EAS1. The primary efficacy assessment, guided by the *treatment-regimen estimand*, will be conducted using FAS.

Unless specified otherwise, safety assessments will be guided by the *safety estimand* comparing safety of each treatment group, irrespective of adherence to study intervention or initiation of antihyperglycemic rescue therapy. The safety analysis will compare tirzepatide groups to placebo during the double-blind period using SS1; the analysis will be conducted during the open-label and follow-up periods using SS2. Selected safety analysis (for example, hypoglycemia) may be conducted excluding data after introducing another antihyperglycemic therapy.

Unless specified otherwise, the analysis of the additional secondary and tertiary/exploratory efficacy endpoints will be aligned to *efficacy estimand analysis* during the double-blind period (0 to 30 weeks) using EAS1 and through Week 52 using EAS2. For selected additional secondary efficacy endpoints (see Section 4.5), *treatment-regimen estimand* analyses will be conducted from baseline to Week 30 using FAS and from baseline to Week 52 using FAS2.

4.1.3. End of Study Participation

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

4.1.4. Treatment Groups

Participants are randomized in 3 treatment arms. During the double-blind study period these intervention arms are Placebo, TZP 5 mg, and TZP 10 mg. During the open-label period there are only 2 interventions: TZP 5 mg and TZP 10 mg.

For the purposes of tables, figures, and listings, during the double-blind study period, the treatment groups will be Placebo, TZP 5 mg, TZP 10 mg, and TZP_ALL (pooled tirzepatide doses). Placebo will be the reference treatment across all statistical reports whenever the treatment effect is compared between TZP 5 mg, TZP 10 mg, TZP_ALL, and placebo.

Up to Week 52 or up to safety follow-up treatment will be expressed in the tables, figures, and listings as TZP_ALL (representing pooled arm of TZP 5 mg/5 mg, TZP 10 mg/10 mg), TZP 10 mg (representing TZP 10 mg/10 mg), TZP 5 mg (representing TZP 5 mg/5 mg), and Placebo/TZP 5 mg (representing placebo/TZP 5 mg). There is no comparison between any of these treatment groups at Week 52.

Within-treatment comparisons between different scheduled visits such as Weeks 30 and 52 for parameters such as HbA1c change (and others) from baseline may be performed.

4.1.5. Analytical Methods

4.1.5.1. Analysis Relative to Efficacy Estimand

The analysis related to the efficacy estimand from double-blind period (baseline to week 30) will be conducted utilizing data in the EAS1 and from double-blind and open-label periods (baseline to week 52) will be conducted using the data in the EAS2.

For double-blind period, the analysis model for continuous measurements to make mean comparisons *over time* among treatment arms relative to the efficacy estimand will be an MMRM. The response variable of the MMRM model will be the change from baseline of the endpoint. Unless specifically provided, the terms in the model are:

- treatment (placebo, tirzepatide 5 mg, and tirzepatide10 mg) as a factor
- visit, as discrete time
- treatment-by-visit interaction
- baseline age group (10 to 14 years, or 15 to <18 years) as a factor
- baseline antihyperglycemic medication (metformin only, or basal insulin only, or metformin and basal insulin) as a factor, and
- baseline of the endpoint as a covariate.

Restricted maximum likelihood will be used to obtain model parameter estimates and the Kenward-Roger option will be used to estimate denominator degrees of freedom. A linear contrast, averaging estimates from the individual tirzepatide doses at 30-week visit, will be used to estimate the treatment effect of the combined tirzepatide arms compared with placebo.

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 order:

- Heterogeneous Toeplitz
- Heterogeneous First Order Autoregressive
- Heterogeneous Compound Symmetry
- Toeplitz
- First Order Autoregressive, and
- Compound Symmetry.

The first covariance structure that converges will be used. For double-blind and open-label periods, a similar MMRM model will be used except with below treatment as a factor:

- treatment (placebo/tirzepatide 5 mg, tirzepatide 5 mg/5 mg, and tirzepatide 10 mg/10 mg)

4.1.5.2. Analytical Approach for Treatment-Regimen Estimand

The analysis relative to the treatment-regimen estimand for continuous endpoints will be conducted using FAS with the aid of an ANCOVA model with missing data imputed with primary multiple imputation method according to Section 4.1.6. The response variable will be the change from baseline of the endpoint measure. The terms of the model will include:

- treatment (placebo, tirzepatide 5 mg, and tirzepatide 10 mg) as a factor
- baseline age group (10 to 14 years, or 15 to <18 years), as a factor
- baseline antihyperglycemic medication (metformin only, or basal insulin only, or metformin and basal insulin) as a factor, and
- baseline of the endpoint as a covariate.

At Week 52, a similar ANCOVA model will be used except with below treatment as a factor: treatment (placebo/tirzepatide 5 mg, tirzepatide 5 mg/5 mg, and tirzepatide 10 mg/10 mg), without any pairwise comparisons between treatment groups.

4.1.5.3. Analytical Approach for Binary Endpoints

For binary endpoints derived from a continuous variable and for both treatment regimen and efficacy estimands, a logistic regression model will be used. For the proportion of participants achieving a specified target value since baseline at the 30-week visit, the following model terms will be included:

- treatment (placebo, tirzepatide 5 mg, and tirzepatide 10 mg),
- baseline age group (10 to 14 years, or 15 to <18 years) as a factor
- baseline antihyperglycemic medication (metformin only, or basal insulin only, or metformin and basal insulin) as a factor, and
- baseline of the endpoint as a covariate.

The unconditional treatment group effect will be assessed by risk difference based on the delta method using formula provided in Ye et al. 2023 (supported by Steingrimsson et al. 2017). The estimated treatment group-specific risk, risk difference, relative risk, p-value and 95% CIs will be presented.
At Week 52 visit, a summary of descriptive statistics will be presented for each estimand.

For missing data imputation: the missing value in the underlying continuous variable will be imputed first (see Section 4.1.6 for details).

4.1.5.4. Analysis Relative to Safety Estimand

The analysis related to the safety estimand from double-blind period (baseline to week 30) will be conducted utilizing data in the SS1, and from double-blind, open-label and follow-up periods will be conducted using the data in the SS2.The analysis models for the continuous endpoints are similar to those specified in Section 4.1.5.1 for MMRM, and in Section 4.1.5.2 for ANCOVA.

The analysis of *safety binary endpoints* or other categorical endpoints will be carried out using Fisher's exact test.

Negative binomial regression model will be used for the treatment comparison of discrete count measures if deemed appropriate. The response variable of the negative binomial model will be number of episodes (for example, the hypoglycemic episodes).

For double-blind period, the model will include:

- treatment (placebo, tirzepatide 5 mg, and tirzepatide 10 mg) as a factor
- baseline hypoglycemia incidence as covariate, and
- with log (exposure in days/365.25) as an offset variable.

For overall study period at Week 52 visit, (including double blind and open-label periods), a summary of descriptive statistics will be presented.

A nonparametric empirical model may be used to analyze the event counts if the event counts are low.

Kaplan-Meier method will be used for estimation of cumulative event-free survival rates over time, and Cox proportional hazards regression analysis will be used to compare hazard rates among treatments. The following terms will be included in the model: treatment, stratification factors (baseline age groups and baseline antihyperglycemic medications).

4.1.6. Missing Data Imputation

ANCOVA analysis in Section 4.1.5.2, in the setting of the *treatment-regimen estimand*, will be conducted after missing primary outcome measures are imputed using J2R imputation (Carpenter et al. 2013) at Week 30. The multiple imputations will be conducted (refer as primary multiple imputation method in this SAP), and the variance estimation will be calculated by combining the within and between imputation variances (Rubin 1987). The J2R imputation assumes the response returns to placebo immediately right after the last observed value at Week 30. Only baseline and endpoint data are used. No intermediate values will be used in the imputation. In the analysis of participants achieving a specified HbA1c target value, relative to the *treatment-*

regimen estimand, missing values at the 30-week visit will be imputed as a continuous outcome using the method above. The imputed values will then be dichotomized into binary outcomes based on cutoff values.

At Week 52, missing data will be imputed using return-to-baseline method under MNAR assumption along with the multiple imputation under MAR assumption. Return-to-baseline imputation assumes missing values will be imputed using baseline assessment.

For analyses aligned to the "efficacy" estimand, missing data will be considered missing at random and an MMRM model, which impute the missing values implicitly under the MAR assumption will be used.

For categorical endpoints, the corresponding continuous variable associated with the missing categorical data will be considered MAR, and multiple imputation assuming the data to be MAR will be performed.

Missing/Invalid Data	**Strategy to Handle Missing/Invalid Data**	**Assumptions for Missing Values**	**Methods to Handle Missing Values**
Data missing at baseline, invalid data collected or missing data after treatment DC due to the leading to invalid measurements ascertained while on treatment, missing data from participants completing the treatment period on the study drug intervention, or missing data after study DC due to inadvertent enrollment.	Hypothetical	MAR	Multiple imputation assuming MAR
Missing data due to any other reason (for example, study DC due to any reason other than inadvertent enrollment).	Treatment policy	MNAR	Jump to Reference at Week 30 and return to baseline at Week 52

Abbreviations: DC = discontinuation; MAR = missing at random; MNAR = missing not at random.

4.1.7. Patient Characteristics

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 and are not limited to age, sex, race, ethnicity, country of enrollment, HbA1c, FSG, duration of T2DM, weight, BMI, and eGFR. Details are specified in Appendix 3.

4.1.8. Concomitant Therapy

The prespecified concomitant medications of interest will be summarized by treatment at randomization. Additionally, medications of interest initiated after randomization will be summarized.

Baseline for concomitant therapy will be defined as the medication and dose that is ongoing at the time of randomization. The concomitant therapies will be mapped using the WHO DRUG dictionary in the clinical trial database and will be further classified using Anatomic-Therapeutic-Chemical codes for reporting purposes. The concomitant medications of interest include the following groups of medication:

- baseline antihypertensive therapy, by type
- baseline lipid lowering therapy, by type
- baseline antihyperglycemic therapy
- rescue therapy due to severe persistent hyperglycemia
- initiation of following medications in double blind period (0 to 30 weeks) and in double-blind and open-label periods (0 to week 52):
    - antidiarrheal medication, and
    - antiemetic medication.
    - antihypertensive medication
    - lipid lowering medication

4.1.8.1. Definition of Rescue Therapy

The definition assumes that a participant has not previously received rescue therapy. Rescue therapy is defined as:

- a participant treated with metformin + study intervention at baseline receives an increased dose of ≥500 mg metformin and/or insulin at any dose for more than 2 weeks
- a participant treated with metformin + basal insulin + study intervention at baseline receives an increased dose of ≥500 mg metformin and/or any increased dose of basal insulin >15% from baseline and/or adding another type of insulin at any dose for more than 2 weeks, or
- a participant treated with basal insulin + study intervention at baseline receives treatment with any increased dose of basal insulin >15% from baseline and/or adding another type of insulin at any dose for more than 2 weeks, OR
- initiation of new antihyperglycemic therapy at any dose for more than 2 weeks.

4.2. Participant Dispositions

Disposition tables will be provided for the double-blind period, and for the entire study. Counts and percentages will be presented by treatment arm using all participants who were screened, randomized, received at least 1 dose of study intervention, or discontinued treatment due to inadvertent enrollment.

The study completion for a participant is defined as the participant completing both the treatment period and the follow-up period, regardless of completion of study treatment.

For all participants in the randomized population, Lilly will provide the reason for discontinuation and the counts and percentages by treatment arm of participants who:

- discontinued study
    - before the end of the double- blind period and

    - before the end of the study.
- discontinued treatment
    - before the end of the double-blind period, and
    - before the end of the study treatment period.

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

4.3. Primary Endpoints Analysis

The primary efficacy measure will be the change in HbA1c from baseline (postbaseline – baseline). Both HbA1c values as well as change from baseline in HbA1c will be summarized by treatment (pooled cohort of tirzepatide doses and placebo) and nominal visit (week).

4.3.1. Primary Analysis Relative to Efficacy Estimand

The primary endpoint analysis relative to the *efficacy estimand* will be carried out using the primary efficacy measure of HbA1c data in the EAS1 from baseline through the 30-week visit with the aid of an MMRM described in Section 4.1.5.1. Missing data are considered as missing at random and will not be explicitly imputed.

The resulting least squares mean estimate of mean change from baseline in HbA1c will be summarized by visit and by study treatment (that is, the pooled cohort of tirzepatide doses and placebo).

A linear contrast, averaging estimates from the individual tirzepatide doses, will be used to estimate the treatment effect of pooled tirzepatide arms compared with placebo.

4.3.2. Main Analytical Approach for Treatment-Regimen Estimand

The primary analysis relative to the *treatment-regimen* estimand will be carried out utilizing the primary efficacy measure of HbA1c data in the FAS at baseline and at the 30-week visit, with missing data imputed (see Section 4.1.6), with the aid of an ANCOVA model described in Section 4.1.5.2.

A linear contrast, averaging estimates from the individual doses, will be used to estimate the treatment effect of the pooled tirzepatide arms compared with placebo. P-values and 2-sided 95% CIs for mean change in HbA1c from baseline to the 30-week visit will be derived and summarized for pooled tirzepatide compared to placebo.

4.3.3. Sensitivity Analyses

A tipping point analysis will be carried out as a means of assessing the influence of multiple imputation on the conclusions of the analyses of the primary efficacy endpoint, and the robustness of the analyses to the multiple imputation.

Typically, in a tipping point analysis, the treatment effect is evaluated several times after adding an incremental shift parameter to the predicted values among subjects with missing data. If the shift parameter needed to overturn the conclusion is so extreme that it is considered clinically implausible, then this indicates robustness to missing data assumptions (Gorst-Rasmussen et al. 2022.).

In the current analysis plan, a "two-way" tipping point analysis will be applied for the sensitivity of the analyses, where different shift parameters are used for tirzepatide and placebo groups.

4.4. Key Secondary Endpoints Analysis

Table GPGV.4.1. Secondary Measures Controlled for Multiplicity.

Relative to the efficacy measure:	Analysis conducted in a manner similar to	Additional Information
Percentage of study participants who achieve ≤6.5% HbA1c reduction from Baseline to Week 30	For both treatment regimen and efficacy estimands: Logistic regression model	Refer to Section 4.1.5.3 for analysis approach detail.
Mean change in BMI SDS (age- and sex-matched) from Baseline to Week 30 using WHO standards Mean percentage change in BMI from Baseline to week 30 Mean change in FSG from Baseline to Week 30 Mean change in HbA1c from baseline to Week 30	For efficacy estimand: MMRM model with treatment group, visit, treatment-by-visit interaction, all stratification factors as fixed effects, and corresponding value at baseline as covariates. For treatment regimen estimand: ANCOVA model with terms of treatment group, stratification factors, and corresponding value at baseline as covariates.	LSM estimates will be plotted by treatment through 30 weeks. Refer to Sections 4.1.5.1 and 4.1.5.2 for analysis approach detail.

Abbreviation: ANCOVA = analysis of covariate; MMRM = mixed model for repeated measures.

Decision will be guided by the 2-sided p-values in each objective subject to type 1 error rate control with details outlined in Section 2.1.

4.5. Additional Secondary and Exploratory Endpoints Analysis

Additional secondary endpoints and exploratory endpoints are not subject to type I error adjustments.

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

Analyses for labs including fasting glucose, HbA1c, and fasting insulin parameters will be performed for both Système International (SI) and conventional (CN) units.

Some parameters may be log-transformed, if necessary.

Table GPGV.4.2 lists all additional secondary and exploratory efficacy endpoints and their analysis relative to treatment regimen and efficacy estimand methods.

Table GPGV.4.2. Additional Secondary and Exploratory Efficacy Measures (Not Controlled for Multiplicity)

Endpoint Relative to the efficacy measure:	Analysis conducted in a manner similar to	Additional Information
Percentage of study participants who achieve HbA1c <7.0%, <5.7 % from Baseline to Week 30 Percentage of study participants who achieve ≥5%, ≥10%, ≥15% BMI reduction from Baseline to Week 30	For both treatment regimen and efficacy estimand: Logistic regression model.	Refer to Section 4.1.5.3 for analysis approach detail.
Percentage of study participants with HbA1c ≤6.5% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia at Week 30 (Exploratory Objectives) Percentage of study participants with HbA1c <7.0% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia at Week 30 (Exploratory Objectives)	For efficacy estimand: Logistic regression model.	Refer to Section 4.1.5.3 for analysis approach detail.
Mean change in BMI from Baseline to Week 30 Mean percentage change in body weight (kg) from Baseline to Week 30 Mean change in body weight (kg) from Baseline to Week 30 Mean change in waist circumference from Baseline to Week 30 Mean change in daily average 6-point SBMG profile from Baseline to Week 30	For efficacy estimand: MMRM model. For treatment regimen estimand: ANCOVA model.	LSM estimates will be plotted by treatment through 30 weeks. Refer to Section 4.1.5.1 and 4.1.5.2 for analysis approach detail.
Percentage of study participants who achieve HbA1c <7.0%, ≤6.5 %, <5.7% from Baseline to Week 52	For both treatment regimen estimand and efficacy estimand, descriptive statistics	Refer to Section 4.1.5.3 for analysis approach detail.

Endpoint Relative to the efficacy measure:	Analysis conducted in a manner similar to	Additional Information
Percentage of study participants who achieve ≥5%, ≥10%, ≥15% BMI reduction from Baseline to Week 52 Percentage of study participants with HbA1c ≤6.5% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia at Week 52 (Exploratory Objectives) Percentage of study participants with HbA1c <7.0% without clinically significant (blood glucose <54 mg/dL) or severe hypoglycemia at Week 52 (Exploratory Objectives)	For efficacy estimand: descriptive statistics	Refer to Section 4.1.5.3 for analysis approach detail.
Mean change in HbA1c from Baseline to Week 52 Mean change in FSG from Baseline to Week 52 Mean change in BMI from Baseline to Week 52 Mean percent change in BMI from Baseline to Week 52 Mean change in BMI SDS (age- and sex-matched) from Baseline to Week 52 using WHO standards	For efficacy estimand: MMRM model. For treatment regimen estimand: ANCOVA model.	LSM estimates will be plotted by treatment through 52 weeks. Refer to Section 4.1.5.1 and 4.1.5.2 for analysis approach detail.
Mean percentage change in body weight (kg) from Baseline to Week 52 Mean change in body weight (kg) from Baseline to Week 52 Mean change in waist circumference from Baseline to Week 52	For efficacy estimand: MMRM model.	LSM estimates will be plotted by treatment through 52 weeks. Refer to Section 4.1.5.1 for analysis approach detail.

Abbreviation: BMI = body mass index; HbA1c = glycated hemoglobin; LSM = least squares mean; MMRM = mixed model for repeated measures; SDS = standard deviation score; SMBG = self-monitored blood glucose; WHO = World Health Organization.

4.5.1. Analysis of Change in Basal Insulin Dose

Unless otherwise specified, all these analyses will be conducted on efficacy estimand. Descriptive summary statistics will be presented for the change from baseline in basal insulin dose (based on total daily dose) for each treatment at each visit from baseline to Week 30 (double blind period) using the EAS1 and from baseline to Week 52 (double-blind and open-label periods) using EAS2.

4.5.2. Insulin Sensitivity and β-Cell Function

Unless otherwise specified, all these exploratory endpoints will be conducted on efficacy estimand. For the double-blind period will be conducted using the EAS1. Change from baseline of the following biomarkers: fasting glucagon, C-peptide, insulin, pro-insulin, adiponectin, homeostatic model assessment for β-cell function, and homeostatic model assessment for insulin resistance will be analyzed separately using an MMRM. The MMRM model will be fitted for each biomarker separately with treatment, time, visit-by-time interaction as fixed effect, subject as a random effect, and corresponding baseline as a covariate. The covariance matrix and other model specifications will be similar to those in Section 4.1.5.1. Typically, a log transformation of the biomarkers measure will be used.

4.6. Safety Analyses

The definitions of AE, SAE, and product complaint can be found in Section 10.3, Appendix 3 of Protocol GPGV.

AEs will be coded from the actual term using MedDRA and reported with PTs and SOC. Selected notable AEs of interest may be reported using HLTs or SMQs.

A TEAE is an untoward medical occurrence that emerges during a defined treatment period, having been absent pretreatment, or worsens relative to the pretreatment state, and does not necessarily have to have a causal relationship with the treatment. 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 with baseline severity.

In general, the baseline of a TEAE is defined as an event prior to first dose of randomized study treatment (tirzepatide or placebo). MedDRA LLT will be used in the treatment-emergent derivation. The maximum severity for each LLT during the baseline period, including ongoing medical history, will be used as baseline severity.

For participants who are randomized to placebo/tirzepatide 5 mg treatment arm at Week 0, the baseline for overall study safety (including all treatment periods and follow-up period) will be defined as:

- for TEAE: an event starts prior to and is ongoing at the time of the first dose of tirzepatide.
- for lab/vitals/ECG categorical analyses: the last non-missing observation collected prior to the time of first dose of tirzepatide.

Unless specified otherwise, safety assessments will be guided by the *safety estimand* setting comparing safety of tirzepatide doses with placebo irrespective of adherence to study intervention or initiation of rescue therapy. Safety analyses for the double-blind period will be conducted using the SS1. For overall study safety, including double-blind period, open-label period, and safety follow-up period, summary statistics will be provided using the SS2.

Percentages of participants with TE abnormal, high, or low laboratory measures at any time will be summarized and compared between treatment groups using Fisher's exact test.

A TE high value is defined as a change from a value less than or equal to the high limit at baseline to a value greater than the high limit at any time during double-blind period, open-label period, and safety follow-up. A TE low result is defined as a change from a value greater than or equal to the low limit at baseline to a value less than the low limit at any time during double-blind period, open-label period, and safety follow-up. High and low laboratory limits will be determined by the central laboratory reference ranges.

Values and change from baseline to postbaseline values of central laboratory measures, vital signs, and selected ECG parameters will be summarized at each scheduled visit. The analysis model to make comparisons among treatment groups, relative to continuous change from baseline values assessed over time can be found in Section 4.1.5.4.

For special safety topics, in case of a smaller number of events, a listing will be provided instead of summary table.

4.6.1. Compliance and Extent of Exposure

Treatment exposure duration is defined as the time from the first dose until the time of the last dose for a given participant. Missing first dosing date will be replaced with the randomization date. Missing last dosing date will be replaced with the date the patient discontinued treatment or completed the treatment period, whichever is earlier.

Duration of follow-up is defined as time in days from date of randomization to date of safety follow-up or last study visit. Duration on study treatment is defined as time in days from date of first dose of study treatment to date of last dose of study treatment plus 7 days.

Overall treatment compliance will be defined as taking at least 75% of the scheduled doses. Compliance will be calculated by taking the number of doses administered, regardless of the actual dose administered, divided by the total number of doses expected to be administered ×100.

A listing of patients who missed ≥3 consecutive doses may be produced. Treatment compliance will be summarized descriptively by treatment using mITT population. Summary of duration of study in double-blinded period, open-label period and, and the follow-up period, and summary of duration on study treatment will be provided by treatment group using mITT.

4.6.2. Adverse Events

Percentages of patients with TEAEs, overall, common, and by maximum severity will be summarized using MedDRA PT. Common TEAEs occurred in ≥5% of patients before rounding in any treatment group will be summarized; maximum severity for the MedDRA PT is the maximum postbaseline severity observed from all associated LLTs mapping to the MedDRA PT. The maximum severity will be determined based on the nonmissing severities. If all severities are missing for the defined postbaseline period of interest, it will be shown as missing in the table. Only counts and percentages will be included for the TEAEs by maximum severity.

Statistical comparisons will be applied at both the SOC and PT levels and risk differences, 95% CIs will be presented. 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.

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

- death
- SAE
- permanent discontinuation of study treatment due to an AE, or
- pregnancy.

An overview of the number and percentage of patients with a TEAE, SAE, and discontinuation from the study or from study treatment due to an AE, as well as the relationship to study treatment will be summarized by treatment from baseline to Week 30 (double-blind period) using SS1, and from baseline to the safety follow-up using SS2. Fisher's exact test will be used to compare the treatment groups for double-blind period.

The number and percentage of patients who prematurely discontinue the study treatment due to an AE will be summarized by treatment using MedDRA PT nested within SOC from baseline to Week 30 and from baseline to the safety follow-up period. Events will be ordered by decreasing frequency within SOC.

4.6.2.1. Deaths

A listing of deaths will be provided using SS2. The listing will include patient identification including the treatment, site number, date of death, age at the time of enrollment, sex, associated AE, time from first dose of study drug to death, time from last dose of study drug to death, cause of death as reported by the investigator.

4.6.2.2. Other Serious Adverse Events

The number and percentage of patients who experienced an SAE during the double-blind period, open-label period, and the follow-up period will be summarized by treatment using MedDRA PT nested within SOC.

Listings of all SAEs will be provided. For the double-blind period, open-label period and the follow-up period, the listing will include treatment, patient identification including the site number, date of the 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, and time from most recent dose to the event (if the patient discontinued study drug prior to the event).

4.6.3. Clinical Laboratory Evaluation

Unless otherwise specified, all the safety lab endpoints will be conducted on safety estimand. Safety analyses for the double-blind period will be conducted using the SS1. For overall study safety, including double-blind period, open-label period, and safety follow-up period, summary statistics will be provided using the SS2.

All laboratory data will be reported in SI units. Selected laboratory measures will also be reported using CN 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 (with respect to the treatment period) and postbaseline values as well as the change from baseline values.

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.

For qualitative laboratory analytes, the number and percentage of patients with abnormal values will be summarized by treatment. All analyses should be conducted from baseline to Week 30, and from baseline to safety follow-up.

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.

For the serum lipid levels including total cholesterol, low density lipoprotein (LDL), high density lipoprotein (HDL), very low-density lipoprotein (VLDL) and triglycerides, an MMRM as described in Section 4.1.5.1 using safety estimand, will be used to fit the changes from baseline at each scheduled postbaseline visits for double-blind period using SS1 and for overall study (including double-blind period, open-label period, and safety follow-up period) using the SS2. Both units will be analyzed using log transformation.

4.6.4. Vital Signs

Unless otherwise specified, vital signs, including physical and developmental measures: height, height SDS using data from WHO, height velocity, height velocity SDS using data from Kelly 2014, and weight SDS using data from CDC 2022 will be analyzed using safety estimand. The analyses for the double-blind period will be conducted using the SS1. For overall study, including double-blind period, open-label period, and safety follow-up period, the analyses will be provided using the SS2.

If multiple records are taken at the same visit, they will be averaged prior to being used for data summaries and analyses.

Height velocity is defined in “cm/year” unit and will be calculated as change in height from baseline to postbaseline scheduled visits at Week 12, Week 30, Week 42 and Week 52divided by the ratio of number of days between both measurements and 365.25. At Visit 3, height velocity will be calculated using the change from screening visit. In addition to height velocity, height velocity SDS will also be presented at same visits.

An MMRM as described in Section 4.1.5.1 will be used to fit the changes from baseline in vital signs at all scheduled postbaseline visits.

Counts and percentages of patients with treatment-emergent abnormal sitting SBP, sitting DBP, and pulse will be presented by treatment. The criteria for identifying patients with treatment-emergent vital sign abnormalities are stated in Table GPGV.4.3.

Table GPGV.4.3. Categorical Criteria for Abnormal Treatment Emergent Blood Pressure and Pulse Measurements (Flynn JT et al 2017)

Parameter	Low	High
Systolic BP (mmHg)	≤85 and decrease from baseline ≥20 (10-11 years old) ≤90 and decrease from baseline ≥20 (≥12 years old)	≥126 and increase from baseline ≥20 (10-11 years old) ≥136 and increase from baseline ≥20 (12-14 years old) ≥140 and increase from baseline ≥20 (≥15 years old)
Diastolic BP (mmHg)	≤50 and decrease from baseline ≥10 (≥10 years old)	≥82 and increase from baseline ≥10 (10-11 years old) ≥86 and increase from baseline ≥10 (12-14 years old) ≥90 and increase from baseline ≥10 (≥15 years old)
Pulse (bpm)	<60 and decrease from baseline ≥25 (10-11 years old) <50 and decrease from baseline ≥15 (≥12 years old)	>140 and increase from baseline ≥25 (10-11 years old) >120 and increase from baseline ≥15 (12-14 years old) >100 and increase from baseline ≥15 (≥15 years old)

Abbreviation: BP = blood pressure; bpm = beats per minute.
High and Low ranges will be derived based on age at the time of assessment.

***Hypertension*:** There may be an incidence of new hypertension, or the worsening of an existing hypertension. Following are the definitions of the 2 categories.

An incidence of new hypertension is defined as one of the following:

- A new AE of hypertension reported, or
- at least 2 consecutive measurements of elevated BP (BP ≥95th percentile for age, sex, and height, or SBP ≥130 mmHg and/or DBP ≥80 mmHg) on separate visits after baseline in an individual with normal BP at baseline.

A worsening of an existing hypertension is defined as one the following:

- AE of hypertension with worsening severity compared to previous MH or AE reported, or

- at least 2 consecutive measurements of elevated BP (BP ≥95th percentile for age, sex, and height, or SBP ≥130 mmHg and/or DBP ≥80 mmHg) on separate visits after baseline in an individual with previous history of hypertension and normal BP at baseline, or
- at least 2 consecutive measurements of elevated BP above baseline (at least 7 mmHg for SBP and 5 mmHg for DBP) on separate visits after baseline in an individual with elevated BP (BP ≥ 95th percentile for age, sex, and height or SBP ≥130 mmHg and/or DBP ≥80 mmHg) at baseline.

The AE of hypertension will be searched using narrow PTs in Hypertension SMQ (20000147). Summary of incidence of new or worsening of hypertension by treatment will be provided from baseline to Week 30 using SS1 and from baseline to Week 52 using SS2.

4.6.5. Electrocardiograms

A listing of the abnormal selected ECG parameters will be produced in all mITT population. Descriptive statistics for the actual measurements and change from baseline, by treatment arm and scheduled visits will be performed for selected ECG parameters from baseline to follow-up using SS2.

QTcF will be used to correct the QT interval using the formulas below:

$$QTcF = QT/RR^{1/3}$$

The analysis of change from baseline in heart rate, PR interval, QRS interval, and QTcF will be conducted using an ANCOVA model with SS1 for double-blind period as specified in Section 4.1.5.4, without imputing any missing data.

Selected thresholds for QTc intervals are shown in Table GPGV.4.4 and will be used to summarize clinically relevant abnormal values for these variables from baseline to Week 30 and from baseline to safety follow-up.

Table GPGV.4.4. Thresholds for ECG Parameters of Interest

ECG Variable (unit)	Threshold
QTcF actual measurement (msec)	>460 (≤15 yrs), >450 (male, ≥16 yrs), >470 (female, ≥16 yrs) >500
QTcF change from Baseline (msec)	>30, >60, >75
ECG heart rate (bpm)	Refer to pulse rate in Table GPGV.4.3.
PR Intervals (msec)	190 (10 to 12 yrs) 210 (13 to 18 yrs)

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; QTcF = Fridericia's corrected QT interval; yrs = years.

Note: PR intervals thresholds were provided by Lambrechts and Fourie (2000).

Threshold will be derived based on age at the time of assessment.

Arrhythmias and cardiac conduction disorders

PTs corresponding to these disorders are listed in the database in Appendix 1.

A summary of the incidence of arrhythmias and of cardiac conduction disorders will be provided with PT by treatment group from baseline to Week 30 using SS1 and from baseline to safety follow-up using SS2.

4.6.6. Special Safety Topics

AESI in this section may be known by alternative equivalent terms. A listing of search criteria for these AESI is provided in Appendix 1.

4.6.6.1. Hypoglycemic Events

Definitions and information on hypoglycemia are provided in Section 8.3.3.1 of Protocol GPGV.

Total hypoglycemia includes any event that meets criteria for documented symptomatic, severe, asymptomatic, or unspecified (that is, missing symptoms or signs, but with a blood glucose, BG, <70 or <54 mg/dL) whether daytime or nocturnal. Those categories that are defined by BG will be analyzed by the BG <70 and <54 mg/dL thresholds separately.

Total hypoglycemia category will have the following 2 subcategories:

- events with BG <70 mg/dL included, and
- events with BG <54 mg/dL included.

To avoid duplicate reporting, all consecutive BG values <70 mg/dL (3.9 mmol/L) occurring within a 1-hour period may be considered as a single hypoglycemic event.

SS1 will be used for the primary analysis for hypoglycemia during Week 0 through Week 30, and SS2 for Week 0 through end of the treatment period (Week 52).

The incidence of hypoglycemic episodes will be summarized for each treatment group for 0 to 30 weeks, and 0 to the end of the treatment period. The summary includes the number of patients and percent of patients reporting hypoglycemic episodes. Fisher's exact test will be used for treatment comparison.

Counts of hypoglycemia or the rate of incidence of hypoglycemia episodes may be analyzed as described in Section 4.1.5.4 with a negative binomial regression model. The results will also be reported for the subset of participants who have baseline basal insulin use. All these analyses will be conducted excluding data after initiation of new antihyperglycemic therapy.

A listing of the individual hypoglycemic episodes will be presented from baseline to the end of the treatment period using SS2. Additional exploratory analyses may be performed if deemed necessary.

4.6.6.2. Severe Persistent Hyperglycemia

Summaries of initiation of rescue therapy (see Section 4.1.8.1) in response to severe, persistent hyperglycemia, will be provided by treatment from baseline to Week 30 using SS1 and from baseline to end of treatment period using SS2. If there are enough episodes, a time-to-first-event analysis for the initiation of rescue therapy may be conducted by treatment using a cox proportional regression model (see Section 4.1.5.4).

For patients without an event, the "time-to-event" will be time, in days, from first dose to end of study participation which is either study discontinuation or end of treatment period. A listing of patients initiating rescue therapy will be provided.

4.6.6.3. Pancreatitis

Treatment-emergent adjudicated-confirmed pancreatitis will be considered as AESI. Summaries of adjudicated and investigator-reported pancreatic events will be provided by treatment from baseline to Week 30 using SS1, and from baseline to safety follow-up using SS2 information about searching criteria for AESI is included in Appendix 1.

Pancreatic enzyme assessment

Observed pancreatic enzyme data (p-amylase and lipase) will be summarized by treatment and nominal visit from baseline to Week 30 using SS1, and from baseline to safety follow-up using SS2. 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 × ULN, >1 × ULN), and post baseline: ≤1 × ULN, and (>1 to ≤3) × ULN, (>3 to ≤5) × ULN, (>5 to ≤10) × ULN, >10 × ULN.

MMRM model as described in Section 4.1.5.4 will be used to analyze each pancreatic enzyme with a log transformed (postbaseline measure/baseline measure) response variable.

4.6.6.4. Thyroid Malignancies, C-Cell Hyperplasia, and Calcitonin

Thyroid malignancies and C-cell hyperplasia will be considered AESI. TE thyroid disease, C-cell hyperplasia, and neoplasms will be identified using predefined MedDRA HLTs of thyroid neoplasms, and PT of C-cell hyperplasia information about searching criteria for AESI is included in Appendix 1. Summary by treatment and PT within HLT, and listings will be provided from baseline to Week 30 using SS1 and from baseline to safety follow-up using SS2.

Observed calcitonin data will be summarized by treatment and nominal visit from baseline to Week 30 using SS1, and from baseline to safety follow-up using SS2. The number and proportion of patients with a maximum postbaseline calcitonin value exceeding the following thresholds will be provided by maximum baseline calcitonin value (≤20 ng/L, >20 ng/L to ≤35 ng/L, >35 ng/L) and post baseline: ≤20 ng/L, >20 to ≤35 ng/L, >35 to ≤50 ng/L, >50 to ≤100 ng/L, >100 ng/L.

4.6.6.5. Diabetic Retinopathy Complications

A summary of new or worsening retinopathy will be provided by treatment through Week 30 using SS1, and through Week 52 using SS2:

- An incidence of new diabetic retinopathy or diabetic maculopathy in either eye is defined as a presence of retinopathy or macular edema in either eye in those without any eye disease at baseline according to dedicated fundoscopic exam form or AE reporting.
- An incidence of worsening diabetic retinopathy or diabetic maculopathy in either eye is defined as a shift to worsening severity of eye disease reported in the follow-up fundoscopic exam form or AE reporting compared to baseline.

- The AE will be searched using predefined PTs to identify events consistent with diabetic retinopathy, diabetic macular edema, and related complications. Detailed search criteria can be found in Appendix 1.

A shift table will be presented using SS2 for double-blind and open-label periods for the results from the dilated fundoscopic exams. The table will show the proportions of participants with shifts in retinopathy status from baseline to most severe postbaseline result using categories defined on the eCRF. These are

- no retinopathy
- mild non-proliferative diabetic retinopathy
- moderate non-proliferative diabetic retinopathy
- severe non-proliferative diabetic retinopathy, and
- proliferative diabetic retinopathy.

Results from visual acuity exam reported on the eCRF will be summarized by treatment and by eyes (that is, any eye, left eye and right eye) for baseline and postbaseline visit (Week 52).

4.6.6.6. Hypersensitivity Events

Serious and severe cases of hypersensitivity by pre-defined SMQ search will be considered as AESI. Hypersensitivity reactions and related information reported will be summarized by treatment. Two main analyses are performed:

- 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, and
- 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.

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 a file with location listed in Appendix 1. Summaries of all hypersensitivity reactions will be generated by PT with decreasing frequency by treatment, using SS1 from baseline to Week 30, and using SS2 from baseline to safety follow-up.

4.6.6.7. Injection Site Reactions

Severe and serious injection site reactions will be considered as AESI. Injection site reactions, incidence, and related information via the “Injection Site Reaction” electronic case report form will be summarized by treatment. Information to be summarized including the timing of the reaction relative to study drug administration, and characteristics of the injection site reaction: erythema, induration, pain, pruritus, and edema.

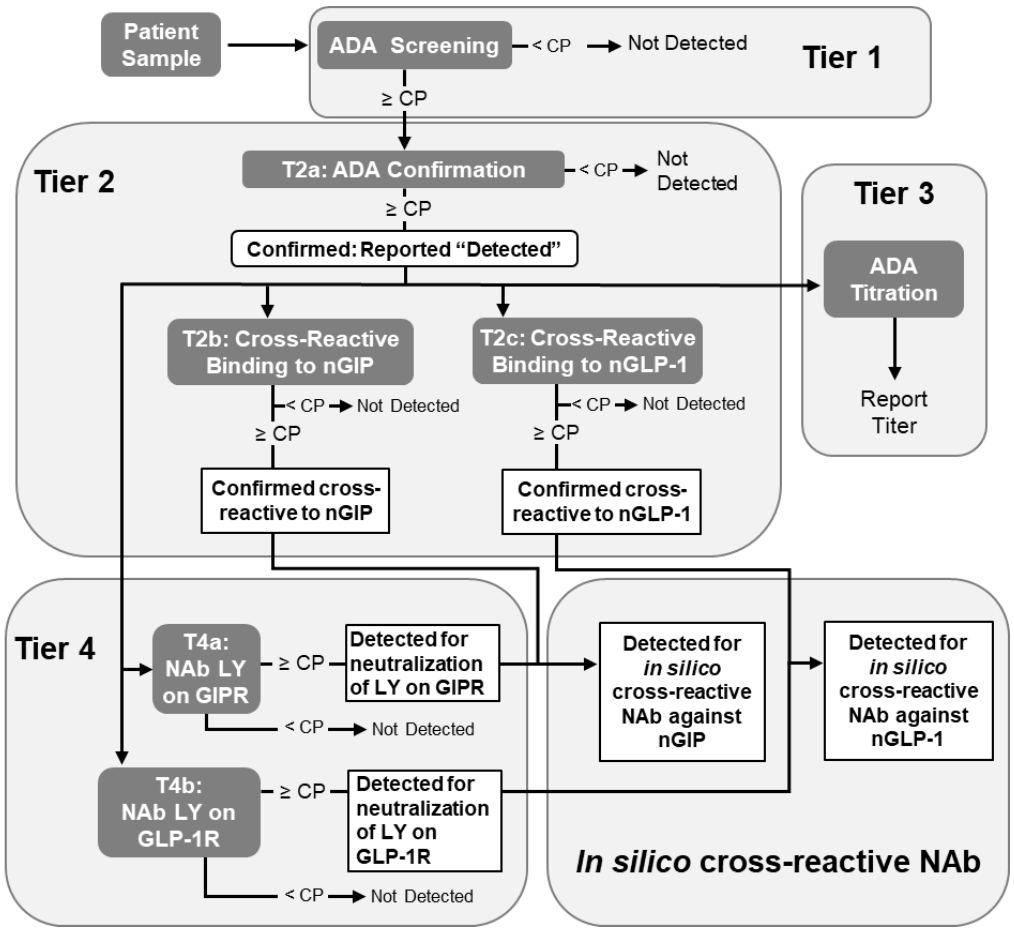
Additionally, potential injection site reactions will be searched by MedDRA HLTs of injection site reactions, administration site reactions, and infusion related reactions. Searching criteria for injection site reaction events can be found in Appendix 1. The PT will be used for summary by treatment within each HLT category in decreasing order of incidence.

4.6.6.8. Immunogenicity

4.6.6.8.1. Definitions of Sample ADA Status

At a high level, an individual sample is potentially examined multiple times, in a hierarchical procedure, to produce a sample ADA assay result and potentially multiple cross-reactive antibody assay results and multiple neutralizing antibody (NAb) assay results.

The cut points used, the drug tolerance of each assay, and the possible values of titers are operating characteristics of the assay. Figure GPGV.4.1 details a flow chart that reflects the multitiered testing approach.



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

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

Table GPGV.4.5 outlines results, as reported from Tier 2a of the multitiered testing approach. Tier 4 results are reported similarly.

Table GPGV.4.5. Sample ADA Assay Results

Sample Laboratory Result	Explanation
Detected	ADA are detected and confirmed.
Not Detected	The raw result as reported from the laboratory indicates not detected. The clinical interpretation of such results depends on other factors (see Table GPGV.4.6).
NO TEST, QNS, etc.	Sample exists but was unevaluable by the assay.

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

It can be the case that the presence of high concentrations of tirzepatide will affect ADA immunoassays. Conversely, high levels of ADA may affect the measurement of tirzepatide concentration. Thus, a tirzepatide drug concentration, assessed from a sample drawn at the same time as the ADA sample, plays a key role in clinical interpretation of a sample when the laboratory result is Not Detected (see Table GPGV.4.6).

Table GPGV.4.6. Sample Clinical ADA Interpretation Results

Sample Clinical Interpretation	Explanation
ADA Present	ADA assay result is Detected
ADA Not Present	ADA assay result is Not Detected and simultaneous drug concentration is at a level that has been demonstrated to not interfere in the ADA detection method (ie, drug concentration is below the assay's drug tolerance level). For participants receiving placebo, drug concentration is not assessed and is assumed to be below the assay's drug tolerance level. If drug concentration was planned but is not available for a treatment-period sample, a Not Detected sample will be declared ADA Not present.
ADA Inconclusive	ADA assay result is Not Detected but drug concentration in the sample is ≥ the assay's drug tolerance level, which may cause interference in the ADA detection method.
ADA Missing	ADA sample not drawn, QNS, not tested, etc., causing there to be no laboratory result reported or the result is reported as "no test."

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

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

Similar terminology to Table GPGV.4.7 applies for each type of cross-reactive and NAb assay. Importantly, each of these are distinct assays and, in general, have different assay operating characteristics.

The following are considered inconclusive for the NAb result:

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

- NAb LY on the GLP-1R: If the NAb result is not detected and the PK concentration is ≥ drug tolerance limit of the NAb LY on the GLP-1R assay.

For cross-reactive NAb interpretations against native GIP and GLP-1, an *in silico* method utilizing results from Tiers 2b and 2c, Tiers 4a and 4b, and tirzepatide concentrations is introduced. The *in silico* method is outlined in Table GPGV.4.7.

Table GPGV.4.7. *In Silico* Classification for Cross-Reactive NAb

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

Abbreviations: ADA = antidrug antibodies; GIPR = glucose-dependent insulinotropic polypeptide receptor; GLP-1R = glucagon-like peptide-1 receptor; N/A = not applicable; NAb = neutralizing antibodies; nGIP = native glucose-dependent insulinotropic polypeptide; nGLP-1 = native glucagon-like peptide-1; Tier 2b = cross-reactive ADA to nGIP; Tier 2c = cross-reactive ADA to nGLP-1; Tier 4a = NAb LY (tirzepatide) on GIPR; Tier 4b = NAb LY (tirzepatide) on GLP-1R.

Note: Only the drug tolerance limits of the Tier 4a and 4b assays are used for in silico classifications as they are lower than the drug tolerance limits of the Tier 2b and 2c assays, respectively.

4.6.6.8.2. Analysis Sets

Table GPGV.4.8 summarizes the immunogenicity analysis sets used in the immunogenicity analysis and for safety and efficacy by ADA status analysis.

Table GPGV.4.8. Immunogenicity Analysis Sets

Analysis Set	Population	Description	Post Baseline Period
Immunogenicity Analysis Set 1	mITT	Refer to SS1 for details (Table GPGV.3.1)	Refer to SS1 for details
Immunogenicity Analysis Set 2	mITT - TZP-treated participants.	Subset of SS2. Participants who had at least one dose of tirzepatide.	1) LY Treatment Period: Data obtained during Double-Blind Period and Open-Label Period, regardless of adherence to study intervention or initiation of rescue antihyperglycemic medication. 2) LY Treatment Period Plus Follow-up Period: Data obtained during Double-Blind Period, Open-Label Period and Safety Follow-up Period, regardless of adherence to study intervention or initiation of rescue antihyperglycemic medication.
Immunogenicity Efficacy Set	mITT - TE ADA evaluable participants	Subset of EAS1.	Refer to EAS1 for details. This analysis set includes only Week 30 completers.

Abbreviations: EAS1 = efficacy analysis set 1; mITT = modified intent-to-treat; SS1 = safety analysis set 1; SS2 = safety analysis set 2; TE ADA = treatment-emergent antidrug antibodies; TZP = tirzepatide.

4.6.6.8.3. Definitions of Immunogenicity Assessment Periods

Immunogenicity Baseline Observations: The baseline period for immunogenicity assessment for each participant includes all observations up to the first dose of study treatment. In instances where multiple baseline observations are collected, to determine participant ADA status the last nonmissing immunogenicity assessment prior to first administration of study drug is used to determine treatment-emergent status (see below).

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

4.6.6.8.4. Definitions of Participant ADA Status

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

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

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

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

TE ADA+ participant: An evaluable participant who had a:

- baseline status of ADA Not Present and at least 1 postbaseline status of ADA Present with titer ≥2 × the MRD of the ADA assay.
- baseline and postbaseline status of ADA Present, with the postbaseline titer being 2 dilutions (4-fold) greater than the baseline titer. That is, the participant has baseline status of ADA Present, with titer 1:B, and at least 1 postbaseline status of ADA Present, with titer 1:P, with P/B ≥4.

As shown in Figure GPGV.4.1, a titer is expected when the ADA assay result is Detected. On the occasion, when the corresponding assay cannot be performed, in which case a titer value will be imputed for the purpose of TE ADA determination. A baseline sample with detected ADA and no titer is imputed to be the MRD (1:10), and a postbaseline sample with ADA detected and no titer is imputed to be 1 dilution above the MRD (1:20).

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

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

For each NAb assay, the following are defined:

NAb+ participant: A participant who is TE ADA+ and has a NAb+ sample in the postbaseline period.

NAb Inconclusive participant: A participant who is TE ADA+ but not NAb+, and all samples that have TE ADA+ titer have a NAb Inconclusive sample result.

Nab negative (NAb−) participant: A participant is who is neither NAb+ nor NAb Inconclusive.

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

4.6.6.8.5. Analyses to be Performed

Table GPGV.4.9 provides the list of planned core and efficacy immunogenicity analyses.

Table GPGV.4.9. List of Core and Efficacy Immunogenicity Analyses

Analysis	Details	Analysis Sets
Summary of TE Tirzepatide Anti-Drug Antibodies Status in Tirzepatide-Treated Participants	Within the participants evaluable for TE ADA, the count and percentage of participants with the following will be summarized: Participants evaluable for TE ADA ADA Present at baseline and NAb Present at baseline, TE ADA+, TE ADA Inconclusive, and TE ADA-, Treatment-induced TE ADA+ and treatment-boosted TE ADA+, TE ADA+ with NAb Present postbaseline, and TE ADA+ without NAb Present postbaseline but having NAb Inconclusive at the follow up visit.	Immunogenicity Analysis Set 1 Immunogenicity Analysis Set 2 (LY Treatment, LY Treatment Plus Follow-up)
Summary of Cross-Reactive and Neutralizing Antibodies from Tirzepatide-Treated Participants with TE ADA	Within the participants evaluable for TE ADA, the count and percentage of participants with the following will be summarized: Participants evaluable for TE ADA ADA, Neutralizing TZP for GIPR, Neutralizing TZP for GLP-1R, GIP Cross-Reactive, In Silico Neutralizing to Native GIP, GLP-1 Cross-Reactive and In Silico Neutralizing to Native GLP-1 Present at baseline, Postbaseline TE ADA+, Neutralizing TZP for GIPR, GIPR Inconclusive, GLP-1R, GLP-1R Inconclusive, GIP Cross-Reactive, GLP-1 Cross-Reactive	Immunogenicity Analysis Set 1 Immunogenicity Analysis Set 2 (LY Treatment, LY Treatment Plus Follow-up)
Summary of TE ADA Persistence for all TE ADA Evaluable Tirzepatide-Treated Participants	Within the participants evaluable for TE ADA, the count and percentages of the following will be summarized: Persistence (Persistent, Potential Persistent, Transient) of TE ADA+ participants, TE ADA- patients	Immunogenicity Analysis Set 2 (LY Treatment, LY Treatment Plus Follow-up)
Maximum Titer Distribution of TE ADA+ Participants	The distribution of the maximum titer values for the TE ADA+ participants will be summarized.	Immunogenicity Analysis Set 2(LY Treatment)
Summary of Time to First TE ADA+ Titer among Tirzepatide-Treated TE ADA Evaluable Participants	A summary table of time to first TE ADA+ titer in TE ADA evaluable participants will be examined by study and treatment group via cumulative categories (≤4 week, ≤12 week, ≤30 week, ≤42 week, ≤52 week, and ≤56 week).	Immunogenicity Analysis Set 2 (LY Treatment)
Summary of Change of Tirzepatide-Treated Participants in HbA1c from Baseline at Primary Visit by TE ADA Status	Analyses on the primary endpoint (change in HbA1c from baseline at Week 30), guided by the efficacy estimand, will be presented visually in the form of boxplots and will include summary statistics for subgroups using TE ADA Status and neutralizing TZP against GIPR and GLP-1R.	Immunogenicity Efficacy Set

Abbreviations: ADA = antidrug antibodies; GIP = glucose-dependent insulinotropic polypeptide; GIP-1 = glucose-dependent insulinotropic polypeptide receptor; GLP-1R = glucagon-like peptide-1 receptor; HbA1c = glycated hemoglobin; Nab = neutralizing antibodies; TE = treatment-emergent; TE ADA = treatment-emergent antidrug antibodies; TZP = tirzepatide.

Safety analysis:

A summary will be provided of the count and percentage of tirzepatide-treated participants experiencing specific TEAEs (Hypersensitivity reactions [see Section 4.6.6.6], Injection site reactions [see Section 4.6.6.7]) by participant TE ADA status (TE ADA+, TE ADA−, and TE ADA Inconclusive). These summaries will utilize data from Immunogenicity Analysis Sets 1. and 2 (LY Treatment).

4.6.6.9. Hepatobiliary Disorders

4.6.6.9.1. Hepatobiliary Events

The detailed search criteria for hepatobiliary disorders can be found in Appendix 1. The counts and percentages of participants with treatment-emergent potentially drug-related hepatobiliary events will be summarized by treatment using the MedDRA PTs contained in the following 2 categories, separately:

- Hepatic events
- Acute gallbladder diseases.

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

4.6.6.9.2. Liver Enzymes

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 measurement from baseline to postbaseline. Baseline categories include ≤1 × ULN, >1 to <3 × ULN, ≥3 to <5 × ULN, and missing. Postbaseline categories include ≤1 × ULN, >1 to <3 × ULN, ≥3 to <5 × ULN, ≥5 to <8 × ULN, ≥8 ×ULN.
- A shift table of maximum-to-maximum aspartate transaminase measurement from baseline to postbaseline. Baseline categories include ≤1 × ULN, >1 × ULN to <3 × ULN, ≥3 to <5 × ULN, and missing. Postbaseline categories include ≤1 × ULN, >1 to <3 × ULN, ≥3 to <5 × ULN, ≥5 to <8 × ULN, ≥8 ×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 from baseline to postbaseline with the following categories: ≤1 × ULN, >1 to <2 × 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 postbaseline period. Planned and unplanned measurements will be included.

An MMRM as described in Section 4.1.5.4 will be used to fit the changes from baseline in these hepatic enzymes (ALT, AST, ALP, total bilirubin and direct bilirubin) for double-blind period

using SS1 and for overall study (double-blind, open-label, and follow-up periods) using SS2. Both units will be analyzed using log transformation.

4.6.6.10. Gastrointestinal Adverse Events

The PTs under the *Gastrointestinal disorders* SOC in MedDRA will be used to identify GI AEs, and only the PTs with serious/severe treatment-emergent cases will be considered as AESIs.

Summaries of all events of TE GI AE, including nausea, vomiting, and diarrhea will be provided by treatment and PT with decreasing frequency.

The time courses of prevalence and incidence (newly occurring episodes) of GI AEs (nausea and/or vomiting and/or diarrhea) will be plotted by treatment and maximum severity.

The maximum severity of TE nausea and/or vomiting and/or diarrhea through the end of the study will be summarized by treatment using SS1 from baseline to Week 30 and using SS2 from baseline to safety follow-up. Time to the onset of nausea, vomiting, and diarrhea will be plotted.

4.6.6.11. Renal Safety

Because severe GI events may lead to dehydration, which could cause a deterioration in renal function including acute renal failure, dehydration events will be analyzed. Acute renal events associated with chronic renal failure exacerbation will also be captured. The detailed search criteria can be found in Appendix 1. The listings of TE renal and TE dehydration events will be provided.

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 and Bedside Schwartz equations with unit mL/min/1.73 m^2, using categories (<30, ≥30 to <45, ≥45 to <60, ≥60 to <90, and ≥90 mL/min/1.73 m^2). A max-to-max shift table of 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).

Descriptive summaries of eGFR by treatment and by nominal visit will be provided for the baseline (with respect to the treatment period) and postbaseline values as well as the change from baseline values. An MMRM as described in Section 4.1.5.4 will be used to fit the changes from baseline in eGFR at all scheduled postbaseline visits using SS1 for double-blind period and using SS2 for overall study (double-blind, open-label, and follow-up periods).

Glomerular hyperfiltration is defined as eGFR >126.8 mL/min/1.73 m^2 using Zappitelli equation as below:

SI units

$$GFR\ (mL/min/1.73\ m^2) = (507.76 \times e^{0.003 \times height})/(CysC^{0.635} \times SCr^{0.547}\ [\mu mol/L])$$

If renal transplant, × 1.165

If spina bifida, × $(SCr^{0.925}\ [\mu mol/L])/40.45$

Conventional units

$$GFR = (43.82 \times e^{0.003 \times height})/(CysC^{0.635} \times SCr^{0.547}\ [mg/dL])$$

If renal transplant, × 1.165

If spina bifida, × $1.57 \times SCr^{0.925}$ (mg/dL)

The glomerular hyperfiltration status shift from baseline to postbaseline will be summarized at Week 30 using SS1 and at Week 52 using SS2.

The MMRM analyses, based on the subset of participants with or without glomerular hyperfiltration at baseline, will be conducted for eGFR with Zappitelli equation using SS1for double-blind period and using SS2 for overall study.

Descriptive summaries of UACR by treatment will be provided for the baseline at Week 30 as well as the change from baseline values. An ANCOVA model, as described in Section 4.1.5.4 without imputing any missing data, will be used to fit the changes from baseline in the logarithm of UACR using SS1 for double-blind period and using SS2 for overall study.

Albuminuria: There might be incidences of new albuminuria or worsening of an existing albuminuria. They are defined as follows:

- Incidence of new albuminuria: A new AE of albuminuria (or equivalent term) reported, or at least 1 measurement of UACR ≥30 mg/g after baseline measured by central laboratory in an individual with normal UACR at baseline.
- Worsening of existing albuminuria: A new AE of albuminuria (or equivalent term) with worsening severity compared to previous MH or AE, or a shift in severity category (mild: UACR 1 to 29, moderate:30 to <300, severe ≥300 mg/g) compared to baseline.

To examine AEs of albuminuria for any suggestion of decrease in renal function, summary of acute renal failure searched by MedDRA SMQ, including incidence of new or worsening of albuminuria will be presented by treatment arm, using SS1 from baseline to Week 30 and using SS2 from baseline to Week 52. The AE of albuminuria will be searched using narrow PTs in *Proteinuria* SMQ (20000220).

4.6.6.12. Metabolic Acidosis, Including Diabetic Ketoacidosis

Metabolic acidosis, including diabetic ketoacidosis will be captured as AESIs. The AE database will be searched using MedDRA PT to identify events consistent with metabolic acidosis, including diabetic ketoacidosis. The incidence of the resulting TEAEs will be summarized by treatment and PT using SS1 from baseline to Week 30 and using SS2 from baseline to safety follow-up (Information regarding searching criteria can be found in Appendix 1).

4.6.6.13. Dyslipidemia

An incidence of new dyslipidemia is defined by 1 of the following 2 conditions:

- a new AE of dyslipidemia reported, or
- at least 2 consecutive values of LDL ≥130 mg/dL or fasting TG ≥150 mg/dL on separate visits after randomization measured by central laboratory in an individual with normal LDL and TG at baseline.

The worsening of an existing dyslipidemia is defined by one of the following conditions:

- AE of dyslipidemia with worsening severity compared to previous MH or AE reported, or
- at least 2 consecutive values of LDL ≥130 mg/dL or fasting TG ≥150 mg/dL on separate visits after randomization measured by central laboratory in an individual with previous history of dyslipidemia and normal LDL and TG at baseline, or
- at least 2 consecutive values of LDL or fasting TG at least 5 mg/dL above baseline on separate visits after randomization measured by central laboratory in an individual with LDL ≥130 mg/dL or TG ≥150 mg/dL at baseline.

The incidence of new or worsening of dyslipidemia will be summarized by treatment, using SS1 from baseline to Week 30, and using SS2 from baseline to Week 52. The AE of dyslipidemia will be searched using narrow PTs in *Dyslipidaemia* SMQ (20000026).

4.6.6.14. Diabetic Neuropathy

There may be an incidence of new diabetic neuropathy reported, or the worsening of an existing one. Incidence of a new diabetic neuropathy, or equivalent term, will be identified by a new AE of diabetic neuropathy reported. A worsening of diabetic neuropathy is an AE of diabetic neuropathy with worsening severity compared to previous MH or AE. The incidence of new or worsening of diabetic neuropathy will be summarized by treatment with PT, using SS1 from baseline to Week 30, and using SS2 from baseline to Week 52 (information regarding searching criteria can be found in Appendix 1).

4.6.6.15. Major Depressive Disorder/Suicidal Ideation

Severe and serious major depressive disorder/suicidal ideation or behavior will be considered as AESI. The AE database will be searched using SMQs to identify events consistent with major depressive disorder or suicidal ideation (information regarding searching criteria can be found in Appendix 1). The incidence of the resulting TEAEs will be summarized by treatment and PT using SS1 from baseline to Week 30 and using SS2 from baseline to safety follow-up.

4.6.7. Product Complaints

A summary of all product complaints, inclusive of device product complaints that lead to an AE and/or SAE will be included by category.

4.7. Other Analyses

4.7.1. Pubertal Progression Evaluation

These analyses will be guided by the safety estimand. A shift table will be presented in those pubertal patients at baseline (Visit 3) to evaluate the change in Tanner Staging from baseline to last post baseline during double-blind period using SS1 and from baseline to last post baseline during overall study (double-blind period and open-label period) using SS2 for male and female groups and each treatment group without pooling. This includes no change and an increase in 1, 2, 3, and 4 levels. Tanner staging will be performed at baseline, Week 16, Week 30, and Week

52. Participants who have reached Tanner level 5 will not be evaluated again. More details about Tanner Staging are in Section 10.9 of Appendix 9 in Protocol GPGV.

4.7.2. Health Outcomes

Item-level missingness is dealt with as per the instrument developer's instruction.

4.7.2.1. EQ-5D-Y

EQ-5D-Y analyses will be guided by Efficacy estimand. The EQ-5D-Y is a standardized generic measure of health status developed by the EuroQoL Group. The EQ-5D-Y consists of the EQ-5D-Y descriptive 3-level system and the EQ VAS. The analysis for the PRO measures (EQ-5D-Y VAS and EQ-5D-Y index scores only if available) will be according to the efficacy estimand with an MMRM as described in Section 4.1.5.1 for the change in PRO score from baseline to Week 52 using the EAS2.

The EQ-5D-Y index score will be analyzed if it is available. Patient responses by EQ-5D-Y domains (mobility; looking after myself; doing usual activities; having pain or discomfort; feeling worried, sad, or unhappy) will also be descriptively examined. The above analyses will be calculated and reported for age 10 to 14 years old, 15 to <18 years old, and all ages combined group. Please refer to the EQ-5D-Y User's Manual for details about scoring are available in the EQ-5D-Y user guide at EuroQoL Research Foundation [https://euroqol.org/publications/user-gfor appropriate descriptive reporting examples, VAS, and how to handle missing data.

4.7.2.2. Pediatric Quality of Life

The 23-item PedsQL 4.0 Generic Core Scales were designed to measure the core dimensions of health as delineated by the (Varni et al. 2019), as well as role (school) functioning. The 2 versions used are for children (ages 8 to 12) and for teens (ages 13 to 18) (Varni et al 2023).

The PEDS-QL 3.2 Diabetes Module is a diabetes-specific health-related quality of life measurement instrument that includes 33 items comprising 5 dimensions for ages 8 and older. Higher scores indicate lower problems.

For the PEDS-QL generic and diabetes scales, transformation of scores are needed as indicated by the scoring manual. Items are scored on a 5-point Likert scale from 0 (Never) to 4 (Almost always). Items are then reversed scored as follows: 0=100, 1=75, 2=50, 3=25, 4=0. The dimension scores and the summary scores are calculated by summing the transformed scores of the items divided by the number of items answered. Moreover, for both the PEDS-QL generic and diabetes scales, if more than 50% of the items in the scale are missing, then the Scale Scores should not be computed. If 50% or more items are completed, mean of the completed items in a scale will be used to impute the missing scores. Refer to the PEDS-QL scoring manual to appropriately score all the dimensions, subscales, and total scores.

The analysis for the PRO measures (PEDS-QL Diabetes 3.2 and PEDS-QL 4.0 [generic scale]) will be analyzed according to efficacy estimand using MMRM, as described in Section 4.1.5.1, for the change in score from baseline to Week 52 using EAS2. Baseline data and each postbaseline visit scores will be calculated and reported for age 10 to 14 years old, 15 to <18 years old, and all age combined group.

The analyses will cover the total scores, all subscale scores, and dimensions.

Dimension Items per for PedsQL 4.0 Generic Core module:

Dimensions	Number of Items
Physical Functioning	8
Emotional Functioning	5
Social Functioning	5
School Functioning	5
Summary Scores	**Items per Summary Score**
Psychosocial (Emotional functioning, social functioning and School functioning)	15
Physical Health (Physical Functioning)	8
Total	23

Dimension Items per for PEDS-QL Diabetes 3.2 module:

Dimensions	Number of Items
Diabetes symptoms	15
Treatment I	5
Treatment II	6
Worry	3
Communication	4

4.7.3. Subgroup Analyses

Efficacy subgroup analyses will be guided by the treatment-regimen estimand in FAS for the primary endpoint at Week 30. Primary multiple imputation method will be used to impute for missing data (Section 4.1.6) consistent with the primary analysis model. The 2-way interaction of treatment-by-subgroup will be evaluated to assess an interaction in the treatment effect with the subgroup levels. Significance will be evaluated at 2-sided alpha of 0.1. The treatment differences between pooled tirzepatide and placebo for each level of subgroup will be presented.

The following subgroup analyses for mean change in HbA1c from baseline to Week 30 will be considered. This list is not necessarily all-inclusive:

- age group (10 to 14 years, or 15 to <18 years)
- race
- ethnicity
- region (US and non-US)
- sex
- duration of diabetes (<median and ≥median)
- baseline BMI (<median and ≥median)
- baseline body weight (<median and ≥median)

- Baseline antihyperglycemic medications use of
    - Metformin only or
    - Basal Insulin only, or
    - Metformin and basal insulin
- baseline HbA1c (≤8.0% or >8.0% [≤64, >64 mmol/mol]).

Safety subgroup analyses will also be conducted based on age group, race, and sex for the most frequently reported TEAEs (occurring in ≥5% patients overall) using SS1.

Other analyses may include analyses of assessments, which are not defined as endpoints, that need to be prespecified and not necessarily be reported in the clinical study report such as, but not limited to, immunogenicity, biomarkers, population PK.

4.7.4. Pharmacokinetics Statistical Inference

All PK/PD analyses will be performed by Lilly or its designee and documented in a separate analysis plan.

4.8. Interim Analysis

No interim analyses of efficacy for early termination are planned for this study. An external DMC will have the responsibility to review unblinded interim analysis results to monitor the safety of the participants in the study.

If an unplanned interim analysis is deemed necessary, the appropriate Lilly medical director, or designee, will be consulted to determine whether it is necessary to amend the protocol.

Unblinding details are specified in a separate unblinding plan document.

4.9. Changes to Protocol-Planned Analyses

Refer to revision history.

5. Sample Size Determination

The statistical power calculation assumes that the evaluation of superiority of pooled tirzepatide 5 mg and 10 mg QW against placebo will be conducted at a 2-sided significance level of 0.05, relative to the primary endpoint (change in HbA1c from baseline at 30 weeks) in an analysis utilizing all data, on or off intervention. Additionally, it assumes a 1.1% greater mean reduction in HbA1c (placebo adjusted) for pooled tirzepatide (5 mg and 10 mg) at Week 30 visit, and a common SD of 1.5% from change in HbA1c at Week 30 visit.

Based on these assumptions, randomizing at least 90 participants in a 1:1:1 ratio to each study arm will provide 90% power to demonstrate superiority of the pooled tirzepatide 5 mg and 10 mg to placebo.

6. Supporting Documentation

6.1. Appendix 1: AESI Search Criteria

The search criteria for each AESI will be documented in an excel spreadsheet stored in Lilly Clinical Users Working Environment. The latest version of MedDRA will be used.

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

Arrhythmias and cardiac conduction disorders
Treatment-emergent arrhythmias, arrhythmias and cardiac conduction disorders will be considered an AESI. The cardiovascular events will include clinically relevant rhythm and conduction disorders. The treatment-emergent arrhythmias and cardiac conduction disorders events will be included using the MedDRA PT contained in any of the following SMQs:

1) **Arrhythmias**:
    a. For symptoms: Arrhythmia related investigations, signs and symptoms SMQ (20000051), narrow and broad terms
    b. For supraventricular arrhythmias: In *Cardiac arrhythmia* SMQ (20000049), under tachyarrhythmia sub SMQ (20000054)
        i. *Supraventricular tachyarrhythmia* SMQ (20000057), broad and narrow terms
        ii. *Tachyarrhythmia terms, nonspecific* SMQ (20000164), narrow terms only; and
        iii. *Ventricular tachyarrhythmia* SMQ (20000058), narrow terms only.
2) Cardiac conduction disorders
    a. *Conduction defects* SMQ (20000056), narrow terms only; and
    b. *Cardiac conduction disorders* HLT (10000032), all PTs.

Hepatic TEAEs
Treatment-emergent potentially drug-related hepatic disorders will be summarized by treatment using the MedDRA PTs contained in any of the following SMQs:

- broad and narrow terms in the *Liver related investigations, signs and symptoms* SMQ (20000008)
- broad and narrow terms in the *Cholestasis and jaundice of hepatic origin* SMQ (20000009)
- broad and narrow terms in the *Hepatitis non-infections* SMQ (20000010)
- broad and narrow terms in the *Hepatic failure, fibrosis and cirrhosis and other liver damage* SMQ (20000013), and
- narrow terms in the *Liver-related coagulation and bleeding disturbances* SMQ (20000015).

Acute gallbladder disease
All events of TEAE biliary colic, cholecystitis, or other suspected events related to gallbladder disease will be summarized by treatment groups by PT with decreasing frequency under the following SMQs:

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

Renal Disorder

- Acute renal failure:
  - Narrow terms in Acute renal failure SMQ (20000003)
- Chronic kidney disease:
  - Narrow terms in Chronic kidney disease SMQ (20000213)

Dehydration events

- Narrow terms in Dehydration SMQ (20000232)

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

C-cell hyperplasia and thyroid malignancies
Thyroid malignancies and C-cell hyperplasia will be identified using MedDRA HLT for *Thyroid neoplasms malignant* and PT for *Thyroid C-cell hyperplasia*.

Hypersensitivity reactions
Analyses are based on the following:

- narrow and algorithm terms in *Anaphylactic reaction* SMQ (20000021)
- narrow terms in *Angioedema* SMQ (20000024)
- narrow terms in *Severe cutaneous adverse reactions* SMQ (20000020), and
- narrow terms in *Hypersensitivity* SMQ (20000214).

For the Anaphylactic reaction SMQ, each term is classified by scope (narrow, broad) and by category (A, B, C, and D). All narrow terms are category A, and all broad terms are category B, C, or D. In addition to the usual narrow and broad searches, the SMQ defines an algorithm to further refine the cases of interest. For immediate analysis, the *Anaphylactic reaction* SMQ algorithm will be included. The algorithm is based upon events that occur within24 hours. The counts and percentages of participants who experienced a TEAE for the following will be analyzed for each of the 2 time periods:

- any narrow term from any 1 of the 4 SMQs indicated above (that is, combined search across narrow of all 4 SMQs), and
- any narrow scope term within each SMQ, separately (that is, narrow SMQ search). For immediate analysis, any term from *Anaphylactic reaction* SMQ algorithm.

Injection site reactions
The ISR AE will be identified using the MedDRA PT in any of the following:

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

Diabetic Retinopathy Complications
The following PT will be used to identify TEAEs of diabetic retinopathy or maculopathy (Section 4.6.6.5):

- Non-proliferative retinopathy
- Arteriosclerotic retinopathy
- Blindness
- Blindness transient
- Blindness unilateral
- Cystoid macular oedema
- Diabetic blindness
- Diabetic eye disease
- Diabetic retinal oedema
- Diabetic retinopathy
- Diabetic uveitis
- Exudative retinopathy
- Eye laser surgery
- Fundoscopy
- Fundoscopy abnormal
- Intra-ocular injection
- Macular detachment
- Macular oedema
- Maculopathy
- Phacotrabeculectomy
- Proliferative vitreoretinopathy
- Retinal collateral vessels
- Retinal cryoablation
- Retinal detachment
- Retinal exudates
- Retinal haemorrhage
- Retinal laser coagulation
- Retinal neovascularization
- Retinal oedema
- Retinal operation
- Retinal thickening
- Retinal vascular disorder
- Retinal vascular occlusion
- Retinopathy
- Retinopathy haemorrhagic

- Retinopathy hypertensive
- Retinopathy proliferative
- Vitrectomy
- Vision blurred
- Visual impairment
- Sudden visual loss
- Visual acuity reduced
- Visual acuity reduced transiently

Diabetic Neuropathy
The following PTs will be used to identify TEAEs of Diabetic Neuropathy (Section 4.6.6.14):

- Acute painful neuropathy of rapid glycaemic control
- Diabetic autonomic neuropathy
- Diabetic mononeuropathy
- Diabetic neuropathy
- Large fibre neuropathy
- Lumbosacral radiculoplexus neuropathy
- Neuropathic arthropathy
- Small fibre neuropathy

Metabolic Acidosis, Including Diabetic Ketoacidosis
The following PTs will be used to identify TEAEs of Metabolic Acidosis, Including Diabetic Ketoacidosis (Section 4.6.6.12):

- Blood ketone body
- Blood ketone body increased
- Blood ketone body present
- Diabetic ketoacidosis
- Diabetic ketoacidotic hyperglycaemic coma
- Diabetic ketosis
- Euglycaemic diabetic ketoacidosis
- Ketoacidosis
- Ketonuria
- Ketosis
- Lactic acidosis
- Urine ketone body

6.2. Appendix 2: Important Protocol Deviations

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

6.3. Appendix 3: Demographic and Baseline Characteristics

Table GPGV.6.1 describes the specific variables and how they will be summarized. The last column specifies variables used for the subgroup analysis described in Section 4.7.3 .

The continuous variables will be summarized using descriptive statistics and the categorical variables will be summarized using frequency counts and percentages. No inferential analysis for the comparability of baseline demographics and characteristics across treatment groups will be performed.

Table GPGV.6.1. Demographics and Baseline Characteristics with Variables for Subgroup Analysis

Variable	Quantitative Summary	Categorical Summary	Subgroup Analysis[a]
Demographics			
Age	Yes	10 to 14 years, 15 to <18 years old	X
Sex	No	Male, Female	X
Ethnicity	No	Hispanic/Latino, Non-Hispanic/Non-Latino	X
Race	No	American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian or other Pacific Islander, White, or Multiple	X
Country	No	By each country	
Geographic region	No	USA, non-USA	X
	No	North America, Europe, Asia, Australia, or Central/South America	
Baseline height (cm)	Yes		
Baseline height SDS	Yes		
Baseline height percentile	Yes		
Baseline waist circumference (cm)	Yes		
Baseline body weight (kg)	Yes	<Median, ≥ Median	X
Baseline BMI (kg/m^2)	Yes	<Median, ≥ Median	X
Baseline BMI SDS	Yes		
Baseline BMI percentile	Yes		
Baseline antihyperglycemic medications			
Baseline antihyperglycemic medications use	No	Metformin only, Basal Insulin only, or Metformin and basal insulin	X
Baseline basal insulin	No	Yes. No	
Baseline Disease characteristics			
Baseline systolic blood pressure (mmHg)	Yes		
Baseline diastolic blood pressure (mmHg)	Yes		
Baseline pulse rate (beats/min)	Yes		
Baseline eGFR CKD-EPI equation, ($mL/min/1.73m^2$)	Yes		
Baseline eGFR bedside Schwartz equations ($mL/min/1.73m^2$)	Yes		
Baseline UACR (g/kg)	Yes		
Baseline triglycerides (mg/dL)	Yes		
Baseline total cholesterol (mg/dL)	Yes		
Baseline VLDL-cholesterol (mg/dL)	Yes		
Baseline LDL-cholesterol (mg/dL)	Yes		

Baseline HDL-cholesterol (mg/dL)	Yes		
Baseline fasting insulin (pmol/L)	Yes		
Baseline HbA1c (%)	Yes	≤8.0% [≤64 mmol/mol], >8.0% [>64 mmol/mol]	X
Baseline HbA1c (mmol/mol)	Yes		
Baseline fasting glucose (mg/dL)	Yes		
Baseline fasting glucose (mmol/L)	Yes		
Duration of T2D (years)	Yes	<Median, ≥ Median	X
Baseline Tanner Stage for Boy Genital	No	G1, G2, G3, G4, G5	
Baseline Tanner Stage for Girl Breast	No	B1, B2, B3, B4, B5	

[a] Subgroup analyses are defined in Section 4.7.3 with more details.

6.4. Appendix 4: Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the CTR requirements.

Analyses provided for the CTR requirements include the following:

- Summary tables of AEs for the double-blind treatment period provided as datasets which will be converted to XML files. For the lead-in period, SAEs are summarized by MedDRA PT. For the treatment period, they are summarized by treatment group and MedDRA PT.
- An AE is considered "serious" whether or not it is a TEAE.
- An AE is considered in the "Other" category if it is both a TEAE and is not serious. For each serious AE and "Other" AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term, and
    - the number of events experienced.
- For each serious AE, these additional terms are provided for EudraCT:
    - the total number of occurrences causally related to treatment.
    - the total number of deaths, and
    - the total number of deaths causally related to treatment.

Demographic table including the following age ranges required by EudraCT: 10 to 17 years.

7. References

Bretz F, Maurer W, Hommel G. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. *Stat Med*. 2011;30(13):1489-1501. https://doi.org/10.1002/sim.3988

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28(4):586-604. https:/doi.org/10.1002.sim.3495

Bretz F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913. https://doi.org/10.1002/bimj.201000239

Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J Biopharm Stat.* 2013;23(6):1352–1371. https://doi.org/10.1080/10543406.2013.834911

[CDC] Centers for Disease Control and Prevention. National Diabetes Statistics Report. 2020. Estimates of diabetes and its burden in the United States. Accessed June 11, 2021. https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

Dabelea D, Mayer-Davis EJ, Saydah S, et al. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *JAMA*. 2014;311(17):1778-1786. https://doi.org/10.1001/jama.2014.3201

EuroQoL Research Foundation. EQ-5D-Y User Guide, 2020. Accessed January 06, 2022. https://euroqol.org/publications/user-guides

Flynn JT, Kaelber DG, Baker-Smith CM, et al. Clinical practice guideline for screening and management of high blood pressure in children and adolescents. *Pediatrics*. 2017;140(3):e20171904. https://doi.org/10.1542/peds.2017-1904

Gorst-Rasmussen A, Tarp-Johansen MJ. Fast tipping point sensitivity analyses in clinical trials with missing continuous outcomes under multiple imputation. *J Biopharm Stat*. 2022;32(6):942-953. https://doi.org/10.1080/10543406.2022.2058525

Kelly A, Winer KK, Kalkwarf H, Oberfield SE, Lappe J, Gilsanz V, Zemel BS. Age-based reference ranges for annual height velocity in US children. *J Clin Endocrinol Metab*. 2014 Jun;99(6):2104-12. doi: 10.1210/jc.2013-4455. Epub 2014 Mar 6. PMID: 24601728; PMCID: PMC4037731.

Mapi Research Trust. Accessed November 11, 2022. https://mapi-trust.org/

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

Varni JW, Delamater AM, Hood KK, et al. Pediatric Quality of Life Inventory (PedsQL) 3.2 Diabetes Module for youth with type 2 diabetes: reliability and validity. *Diabet Med*. 2019;36(4):465-472.

Varni JW, SCALING AND SCORING for the Acute and Standard versions OF THE Pediatric Quality of Life Inventory™ PedsQL™, access 06January2025, (https://www.pedsql.org/PedsQL-Scoring.pdf)

WHO resources page. World Health Organization website. Accessed November 10, 2022. https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age

Ye T, Bannick M, Yi Y, Shao J. 2023. Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. arXiv preprint arXiv:2302.10404.

Ye T, Shao J, Yi Y, Zhao Q. Toward better practice of covariate adjustment in analyzing randomized clinical trials. *J Am Stat Assoc*. 2022. https://doi.org/10.1080/01621459.2022.2049278

Signature Page for VV-CLIN-140059 v2.0

Approval	PPD 15-Jan-2025 15:35:49 GMT+0000

Signature Page for VV-CLIN-140059 v2.0

Signature Page for VV-CLIN-140059 v3.0

Approval	PPD 15-Jan-2025 19:28:32 GMT+0000

Signature Page for VV-CLIN-140059 v3.0