Statistical Analysis Plan J2A-MC-GZGP Version 3.0

A Phase 3, Randomized, Double-Blind Study to Investigate the Efficacy and Safety of Once-Daily Oral Orforglipron Compared With Placebo in Adult Participants With Obesity or Overweight With Weight-Related Comorbidities (ATTAIN-1)

NCT05869903

Approval Date: 25 JUN 2025

Title Page

Protocol Title: A Phase 3, Randomized, Double-Blind Study to Investigate the Efficacy and Safety of Once-Daily Oral LY3502970 Compared with Placebo in Adult Participants with Obesity or Overweight with Weight-Related Comorbidities (ATTAIN-1)

Protocol Number: J2A-MC-GZGP

Compound Number: LY3502970 (orforglipron)

Short Title: GZGP SAP

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

Registry	ID
IND:	156143
EU trial number	2022-502839-19-00

Document ID: VV-CLIN-117409

Approved on 25 Jun 2025 GMT

Table of Contents

Table of Contents

Tables **Page**

Approved on 25 Jun 2025 GMT

Table of Contents

Figure **Page**

Version History

The statistical analysis plan (SAP) Version 1 for Study J2A-MC-GZGP (GZGP) was based on Protocol GZGP amendment (c), dated 23 October 2023.

GZGP SAP Version 2 included updates based on regulatory feedback. The major updates in the Version 2 amendment are listed in the table below.

GZGP SAP Version 3 was updated based on Protocol GZGP amendment (d), dated 30 April 2025, and regulatory feedback. The major updates in this amendment are listed in the table below.

SAP Version History Summary

SAP Version	Approval Date	Section	Change	Rationale
1	January 24, 2024		Not Applicable	Original version
2	November 6, 2024	1.1	Added prohibited glycemic therapy to Week 176 estimands	Clarification
		3	Removed note regarding no inferential comparisons for the summary of disposition, demographics, historical illness, and preexisting conditions	Clarification
		4.1.2.1.2	Further clarified the definition of Scenario 5 Removed the direct imputation method Updated the imputation method for Scenario 5B Added footnote in Table GZGP.4.4 for protocol deviation, inadvertent enrolment, and emergency unblinding	The direct imputation method is not fully aligned with the primary imputation strategy. Used a different method to handle Scenario 5B, based on simulation results from Liu et al. (2024) Clarified the distinction of imputing discontinuation due to

SAP Version	Approval Date	Section	Change	Rationale
				protocol deviation, inadvertent enrolment
		4.1.2.1.4	Updated the penalty range for the MI-TP analysis to -15% to 15%	Updated due to magnitude of projected weight loss
		4.1.2.2.3	Added a censoring condition for date of initiation of prohibited medications	To align with the defined estimand
		4.1.2.3	Updated MMRM model for safety analysis Clarified the definition of start time for different analyses	To be consistent with the PSAP
		4.4	Updated the derivation for the appetite VAS item and overall scores	To align with CRFs and the literature
		4.4	Added PRO analyses for the extension	Clarification
		4.6.6	Modified Table GZGP.4.11	To be consistent with the PSAP
		4.7.1	Updated the testing of subgroup effects. Added details for the Bayesian shrinkage method	Clarification
		Table GZGP.6.1	Updated subgroup analyses	Clarification
		Table GZGP.6.5	Added anti-hyperglycemic prohibited meds	To align with protocol
		6.9	Updated the statistical analysis for China	Clarification

SAP Version	Approval Date	Section	Change	Rationale
		6.10	Updated the statistical analysis for Japan	Clarification
3	See date on Page 1	1.1	Moved 6 mg endpoint "percentage of participants who achieve a body weight reduction of ≥20%" from key secondary list to additional secondary list, and updated hypotheses list accordingly	Limited clinical meaningfulness in anticipated percentage of participants achieving ≥20% weight reduction from 6 mg.
		1.1	Added Key secondary objective "Delayed progression to T2D at 190 weeks"	To align with protocol (d)
		1.1	Updated language for PGI endpoints	Clarification
		1.1	Added exploratory objectives for participants with prediabetes at randomization at 176 and 190 weeks	Clarification
		1.1	Remove prohibited glycemic therapy as an intercurrent event for the delayed progression to T2D endpoint	Due to rare use and minimal impact given the non-T2D study population
		1.1	Added safety estimand for the newly added key 2nd objective "Delayed progression to T2D at 190 weeks"	To align with protocol (d)

SAP Version	Approval Date	Section	Change	Rationale
		1.2	Added description for primary and final database lock	Clarification
		2.1	Updated the graphical testing scheme with the final version	Clarification
		3	Minor updates on data points sets	Clarification
		4	Replaced country with region in the statistical analysis models	To improve model convergence
		4.1.2.1.1	For method for continuous variables, added description for obtaining estimates of the pooled dose vs. placebo	Clarification
		4.1.2.1.2	Clarified the imputation strategy for all scenarios Updated study discontinuation reason categories for imputation purpose	Clarification
		4.1.2.1.5	For the method for binary variables, added interaction terms in logistic regression, and clarified the statistical testing for risk difference is the primary approach for treatment comparisons	Consistency with other analyses Clarification
		4.1.2.3	Replaced negative binomial model with empirical method for	The approach aligns with the limited events.

SAP Version	Approval Date	Section	Change	Rationale
			analysis of rate of hypoglycemia	
		4.4	In Table GZGP.4.5, modified and added variables	Clarification and to align with study endpoints
		4.4	In Table GZGP.4.6, modified and added additional analyses	Clarification and to align with study objectives
		4.6.6	Modified Table GZGP.4.11	To be consistent with the PSAP
		4.7.1	Added details for the Bayesian shrinkage method for subgroup analysis	Clarification
		6.7 (Appendix 7)	Modified parameter list and provide additional details on analyses method	Clarification
		Throughout SAP	Minor changes and reorganization	For clarity, no change to analysis methodologies, so not detailed

Abbreviations: MI-TP = multiple imputation tipping point; MMRM = mixed model for repeated measures; PGI = Patient Global Impression; PRO = Patient-Reported Outcome; PSAP = program safety analysis plan; SAP = statistical analysis plan; T2D = type 2 diabetes; vs = versus.

1. Introduction

This SAP is intended to describe the analyses of primary and secondary objectives, as well as safety assessments for Study GZGP. Sensitivity and supplementary analyses intended to support the primary and key secondary objectives are also included. Additional exploratory analyses, if needed, may be included in a separate document.

Pharmacokinetic/pharmacodynamic (PK/PD) analyses will be described in a separate document.

Changes to the protocol-planned analyses, if any, are described in Section 4.9.

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary Objective	
To demonstrate that orforglipron 6 mg, 12 mg, and/or 36 mg QD is superior to placebo for • body weight.	From baseline to Week 72 • mean percent change in body weight.
Key Secondary Objectives (controlled for Type 1 error)	
To demonstrate that orforglipron 12 mg, and/or 36 mg QD is superior to placebo in change from baseline for the following: • body weight • waist circumference.	From baseline to Week 72 • percentage of participants who achieve a body weight reduction of: ○ ≥5% ○ ≥10% ○ ≥15%, and ○ ≥20% mean change in waist circumference (cm)
To demonstrate that orforglipron 6 mg QD is superior to placebo in change from baseline for the following: • body weight • waist circumference.	From baseline to Week 72 • percentage of participants who achieve a body weight reduction of: ○ ≥5% ○ ≥10%, and ○ ≥15% mean change in waist circumference (cm)
To demonstrate that orforglipron (6 mg, 12 mg, and 36 mg QD pooled doses) is superior to placebo in change from baseline for the following: • systolic blood pressure, and • lipid parameters.	From baseline to Week 72 • mean change in systolic blood pressure (mmHg) • mean percent change in fasting ○ non-HDL cholesterol, and

	○ triglycerides.
Key Secondary Objectives at 176 weeks for participants with prediabetes at randomization (controlled for Type 1 error), pooled dose analysis	
To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and 36 mg QD pooled doses) is superior to placebo at 176 weeks:	From baseline to Week 176
• body weight.	• mean percent change in body weight
Key Secondary Objectives at 176 or 190 weeks for participants with prediabetes at randomization (controlled for Type 1 error), pooled dose analysis	
To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and 36 mg QD pooled doses) is superior to placebo for:	
• Delayed progression to T2D at 176 weeks • Delayed progression to T2D at 190 weeks.	• Time to onset of T2D during 176-week treatment period • Time to onset of T2D during entire study including post-treatment follow up period
Additional Secondary Objectives	
To demonstrate that orforglipron 6 mg, 12 mg, and/or 36 mg QD is superior to placebo in change from baseline for the following	From baseline to Week 72
• body weight	• mean change in absolute body weight (kg) • mean change in body mass index (kg/m^2)
• glycemic control	• mean change in HbA1c (%) • mean change in fasting glucose (mg/dL)
• fasting insulin.	• mean percent change in fasting insulin.
To demonstrate that orforglipron 6 mg QD is superior to placebo in change from baseline for the following:	From baseline to Week 72

• body weight	• percentage of participants who achieve a body weight reduction of: ○ ≥20%

To demonstrate that orforglipron (6 mg, 12 mg and 36 mg QD pooled doses) is superior to placebo in change from baseline for the following:
• diastolic blood presssure
• lipid parameters, and
• patient-reported outcomes
	From baseline to Week 72
	• mean change in diastolic blood pressure (mmHg)
	• mean percent change from baseline in fasting
	  ○ total cholesterol
	  ○ LDL cholesterol, and
	  ○ HDL cholesterol
	• mean change in SF-36v2 acute form domain scores
	• mean change in EQ-5D-5L health state utilities and VAS
	• mean change in IWQOL-Lite-CT Physical Function, Physical, and Psychosocial composite scores, and total score
	• proportion of participants with improved categorical shift in PGI-S physical function due to weight, and
	• proportion of participants with improvement in PGI-C physical function due to weight

To describe the safety of orforglipron as compared to placebo	Summary of safety data, including number and incidence of • treatment-emergent adverse events
Additional Secondary Objectives at 72 weeks in participants with prediabetes at randomization	
To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and/or 36 mg QD) is superior to placebo in change from baseline for • glycemic control.	From baseline to Week 72 • percentage of participants achieving normoglycemia
Additional Secondary objectives at 176 weeks in participants with prediabetes at randomization	

To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and/or 36 mg QD) is superior to placebo for the following at 176 weeks: • body weight • glycemic control	From baseline to Week 176 • percentage of study participants who achieve ≥5% body weight reduction • mean percent change in body weight • mean change in HbA1c (%) • mean change in fasting glucose (mg/dL), and • percentage of participants achieving normoglycemia
Tertiary Objectives	
To characterize the population PK of orforglipron and explore the relationships between orforglipron concentration and efficacy, safety, and tolerability measures.	• population PK and PD parameters
Exploratory Objectives	
To determine the effects of orforglipron 6 mg, 12 mg, 36 mg QD, and pooled orforglipron doses (6, 12, and 36 mg) on appetite sensations and desire for specific foods • appetite VAS.	From baseline to Week 24 and 72 • mean change in item appetite VAS scores, and • mean change in overall appetite VAS score. From baseline and Week 72 to Week 74 in participants without prediabetes at randomization • mean change in item appetite VAS scores, and • mean change in overall appetite VAS score. From baseline to Week 176 in participants with prediabetes at randomization • mean change in item appetite VAS

	scores, and • mean change in overall appetite VAS score.
To assess the effects of orforglipron 6 mg, 12 mg, 36 mg QD, and pooled orforglipron doses (6, 12, and 36 mg) on the psychological impact of living in an environment with an abundance of palatable foods • PFS.	From baseline to Week 24 and 72 • mean change in domain PFS scores, and • mean change in overall PFS score. From baseline and Week 72 to Week 74 in participants without prediabetes at randomization • mean change in domain PFS scores, and • mean change in overall PFS score. From baseline to Week 176 in participants with prediabetes at randomization • mean change in domain PFS scores, and • mean change in overall PFS score.
To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and/or 36 mg QD) is superior to placebo for the following at 176 weeks: • body weight • waist circumference • fasting insulin	From baseline to Week 176 • percentage of participants who achieve a body weight reduction of: ○ ≥10% ○ ≥15%, and ○ ≥20% • mean change in absolute body weight (kg) • mean change in BMI (kg/m^2) • mean change in waist circumference (cm) • mean percent change in fasting insulin
To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and 36 mg QD pooled doses) is superior to placebo at 176 weeks:	From baseline to Week 176

• blood pressure • lipid parameters	• mean change in SBP (mm Hg) • mean change in DBP (mm Hg) • mean percent change in fasting o non-HDL cholesterol o triglycerides o total cholesterol o LDL cholesterol o HDL cholesterol
To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and/or 36 mg QD) is superior to placebo for the following at 190 weeks: • body weight • waist circumference • glycemic control	From baseline to Week 190 • mean percent change in body weight • mean change in absolute body weight (kg) • mean change in waist circumference (cm) • mean change in HbA1c (%) • mean change in fasting glucose (mg/dL), and • percentage of participants achieving normoglycemia
To demonstrate in participants with prediabetes at randomization that orforglipron (6 mg, 12 mg, and 36 mg QD pooled doses) is superior to placebo at 190 weeks: • blood pressure • lipid parameters	From baseline to Week 190 • mean change in SBP (mmHg) • mean change in DBP (mmHg) • mean percent change in fasting o non-HDL cholesterol o triglycerides o total cholesterol o LDL cholesterol o HDL cholesterol

Abbreviations: DBP = diastolic blood pressure; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite Clinical Trials Version; LDL = low-density lipoprotein; PD = pharmacodynamics; PFS = Power of Food Scale; PGI-C = Patient Global Impression-Change; PGI-S = Patient Global Impression-Severity; PK = pharmacokinetics; QD = once daily; SBP = systolic blood pressure; SF-36v2 Acute Form = Short Form 36 Health Survey Acute, version 2; T2D = type 2 diabetes; VAS = visual analog scale.

Estimands

There will be 2 estimands for the primary objective planned in Study GZGP. These estimands address intercurrent events (ICEs) using either the treatment policy strategy or the hypothetical strategy (ICH 2021), respectively.

Estimands for weight-related, blood pressure, and lipid parameters

The below estimands will be applied to weight-related, blood pressure, and lipid parameters, using percent change in body weight at Week 72 visit as an example. For other weight-related, blood pressure, and lipid parameters, replace percent change in body weight by the endpoints to be analyzed. For endpoints at other time points, replace Week 72 with the appropriate timepoint.

Treatment regimen estimand

The treatment regimen estimand will be the primary estimand.

The clinical question of interest: What is the treatment difference in the percent change in body weight from baseline to 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals who meet eligibility criteria regardless of adherence to study intervention or initiation of prohibited weight management treatments?

Treatment policy strategy

The occurrence of the ICE is considered irrelevant in defining the treatment effect of interest; the values for the variable of interest are used regardless of whether the ICE occurs.

The treatment regimen estimand is described by the following attributes

1) ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
2) ***Endpoint:*** Percent change in body weight from baseline to 72 weeks.
3) ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications regardless of adherence to study intervention or initiation of prohibited weight management treatments.
4) ***Intercurrent events:*** No ICEs since treatment adherence and the initiation of prohibited weight management treatments are part of the treatment condition.

5) ***Population-level summary and treatment effect of interest:*** The difference in mean percent change from baseline in body weight to Week 72 between orforglipron and placebo.

Rationale for the estimand: This estimand aims to evaluate the efficacy of orforglipron that reflects the real-life behavior of the target population.

Efficacy estimand

The clinical question of interest: What is the treatment difference in the percent change in body weight from baseline to 72 weeks between orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals who meet the eligibility criteria if they would remain on their randomly assigned treatment for 72 weeks and would not initiate prohibited weight management treatments?

Hypothetical strategy

A scenario is envisaged in which the ICE would not occur. The value of the variable to reflect the clinical question of interest is the value that the variable would have taken in the hypothetical scenario defined.

The efficacy estimand is described by the following attributes:

6) ***Population:*** Individuals who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
7) ***Endpoint:*** Percent change in body weight from baseline to 72 weeks.
8) ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications.
9) ***Intercurrent events:*** ICEs include permanent discontinuation of study intervention and initiation of prohibited weight management treatments, which is handled by the hypothetical strategy. The potential outcome of interest is the response in the efficacy measurement if participants would remain on their randomly assigned treatment for 72 weeks and would not initiate prohibited weight management treatments. Dose modification and interruption will not be considered as ICEs since they are part of the treatment condition.
10) ***Population-level summary and treatment effect of interest:*** The difference in mean percent change in body weight from baseline to Week 72 between orforglipron and placebo.

Rationale for the estimand: This estimand aims to evaluate the efficacy of orforglipron under the ideal condition that all participants adhere to their randomly assigned treatment without being confounded by the initiation of other weight management treatments.

Estimands for delayed progression to type 2 diabetes

The below estimands will be applied to time to onset of type 2 diabetes (T2D) for participants with prediabetes at randomization.

Treatment regimen estimand

The treatment regimen estimand will be the primary estimand.

The clinical question of interest: What is the hazard ratio (HR) of developing T2D from randomization up to 176 weeks between pooled orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals with prediabetes who meet the eligibility criteria regardless of adherence to study intervention or initiation of prohibited weight management treatments?

The treatment regimen estimand is described by the following attributes

11) ***Population:*** Individuals with prediabetes who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
12) ***Endpoints:*** Time from randomization to onset of T2D up to 176 weeks.
13) ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications regardless of adherence to study intervention or initiation of prohibited weight management treatments.
14) ***Intercurrent events:*** No ICEs are defined as treatment adherence and the initiation of prohibited weight management treatments are a part of the treatment condition.
15) ***Population-level summary and treatment effect of interest:*** The HR between pooled orforglipron and placebo for the time from randomization to onset of T2D.

Rationale for the estimand: The estimand aims to evaluate the efficacy of orforglipron that reflects the real-life behavior of the target population.

Efficacy estimand

The clinical question of interest: What is the HR of developing T2D from randomization up to 176 weeks between pooled orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals with prediabetes who meet the eligibility criteria if they would remain on their randomly assigned treatment for 176 weeks and would not initiate prohibited weight management treatments?

The efficacy estimand is described by the following attributes

16) ***Population:*** Individuals with prediabetes who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
17) ***Endpoint:*** Time from randomization to onset of T2D up to 176 weeks.
18) ***Treatment condition:*** The randomized treatment with allowance for potential dose interruptions and modifications.
19) ***Intercurrent events:*** ICEs include permanent discontinuation of study intervention and initiation of prohibited weight management treatments, which are handled by the

hypothetical strategy. The potential outcome of interest is the response in the efficacy measurement if participants would remain on their randomly assigned study intervention for 176 weeks and would not initiate prohibited weight management treatments. Dose modification and interruption will not be considered as ICEs since they are a part of treatment condition.

20) ***Population-level summary and treatment effect of interest:*** The HR between pooled orforglipron and placebo for the time from randomization to onset of T2D.

Rationale for the estimand: This estimand aims to evaluate the efficacy of orforglipron under the ideal condition that all participants would adhere to the randomly assigned study intervention without being confounded by the initiation of prohibited weight management treatments.

Safety estimand

The clinical question of interest: What is the hazard ratio (HR) of developing T2D from randomization up to 190 weeks (including a 14-week post-treatment follow-up period) between pooled orforglipron 6 mg, 12 mg, and 36 mg versus placebo, as an adjunct to healthy diet and physical activity, in individuals with prediabetes who meet the eligibility criteria regardless of adherence to study intervention or initiation of prohibited weight management treatments?

The safety estimand is described by the following attributes

21) ***Population:*** Individuals with prediabetes who meet the eligibility criteria. This represents the target population identified through study inclusion/exclusion criteria (ICH E9). For the analysis, intent-to-treatment principle will be applied. Details are defined in Section 3.
22) ***Endpoints:*** Time from randomization to onset of T2D up to 190 weeks.
23) ***Treatment condition:*** The randomized treatment including a 14-week posttreatment follow-up period with allowance for potential dose interruptions and modifications regardless of adherence to study intervention or initiation of prohibited weight management treatments.
24) ***Intercurrent events:*** No ICEs are defined as treatment adherence and the initiation of prohibited weight management treatments are a part of the treatment condition.
25) ***Population-level summary and treatment effect of interest:*** The HR between pooled orforglipron and placebo for the time from randomization to onset of T2D.

Rationale for the estimand: The estimand aims to evaluate the efficacy of orforglipron that reflects the real-life behavior of the target population including both the treatment and the posttreatment follow-up period.

1.2. Study Design

Study GZGP is a Phase 3, multicenter, randomized, parallel-arm, double-blind, placebo-controlled study. Study GZGP will investigate the safety and efficacy of treatment with daily oral doses of orforglipron (6 mg, 12 mg, or 36 mg), compared with placebo in participants without T2D with either have obesity (body mass index [BMI] 30 kg/m^2 or greater), or overweight (BMI 27 kg/m^2 or greater) with the presence of at least 1 weight-related comorbidity

(for example, hypertension, dyslipidemia, obstructive sleep apnea, or cardiovascular [CV] disease). Eligible participants will be assigned to either 72 or 176 weeks of treatment based upon baseline prediabetes status (no prediabetes and prediabetes, respectively).

Study GZGP participants will be randomly assigned in a 3:3:3:4 ratio to receive a daily dose of orforglipron (6 mg, 12 mg, or 36 mg) or placebo. An upper limit of 70% enrollment of females will be used to ensure a sufficiently large sample of males.

Study GZGP includes a

- screening period: 3 weeks
- treatment period:
  - dose escalation period: 20 weeks
  - maintenance dose period
    - no prediabetes: 52 weeks
    - prediabetes: 156 weeks total (including initial 52-week treatment period and 104-week additional prediabetes treatment period), and
- posttreatment follow-up period:
  - no prediabetes, or participants discontinuing study intervention during the first 72 weeks: 2 weeks
  - prediabetes: 14 weeks

The planned duration of treatment for the primary endpoint at 72 weeks allows for at least a 52-week treatment period at the randomly assigned dose (6 mg, 12 mg, or 36 mg). The effects of study intervention cessation will be assessed at the 2-week posttreatment follow-up visit (Week 74).

Primary database lock will occur after:

- all randomized participants with normoglycemia at baseline either complete Visit 801 or discontinue prior to Visit 801.
- all randomized participants with prediabetes at baseline either complete the 72-Week visit or discontinue the study prior to the 72-week visit.

The primary objective of Study GZGP will be assessed following the primary database lock.

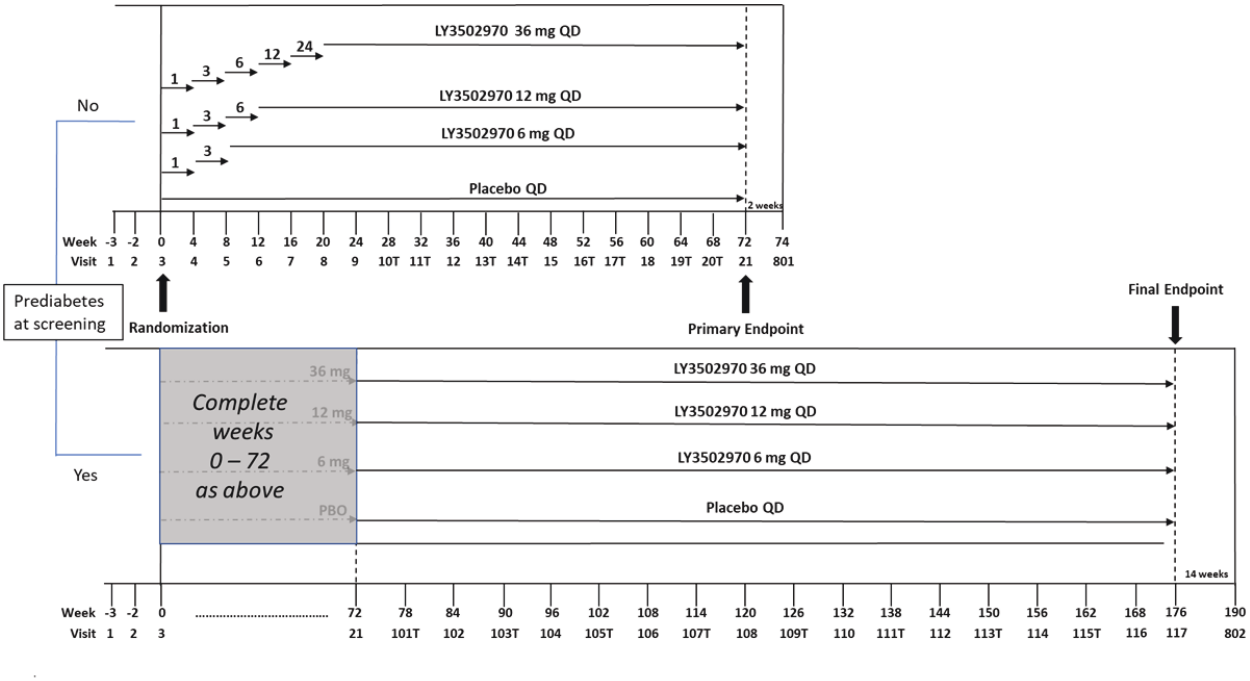
To obtain additional information regarding the time to new onset of T2D while taking orforglipron, participants with prediabetes diagnosed at the beginning of Study GZGP (who are at increased risk of diabetes), will be treated and observed for an additional 2 years. For participants with prediabetes at randomization, the effects of study intervention cessation will be assessed at a 14-week posttreatment follow-up visit (Week 190).

The final database lock will occur after participants with prediabetes at randomization have completed a 176-week treatment period and a 14-week follow-up period or discontinued the study early.

The Study GZGP schema is shown in Figure GZGP.1.1.

The randomization will be stratified by:

- prediabetes status (yes, no)
- sex (female, male), and
- country.



Abbreviations: PBO = placebo, QD = once daily; T= telehealth visit.

Figure GZGP.1.1. Illustration of study design for Clinical Protocol J2A-MC-GZGP.

2. Statistical Hypotheses

The null hypotheses corresponding to the primary objective are as follows

- $H_{1,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the mean percent change from baseline in body weight at Week 72.
- $H_{2,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the mean percent change from baseline in body weight at Week 72.
- $H_{3,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the mean percent change from baseline in body weight at Week 72.

The null hypotheses corresponding to the key secondary objectives are as follows

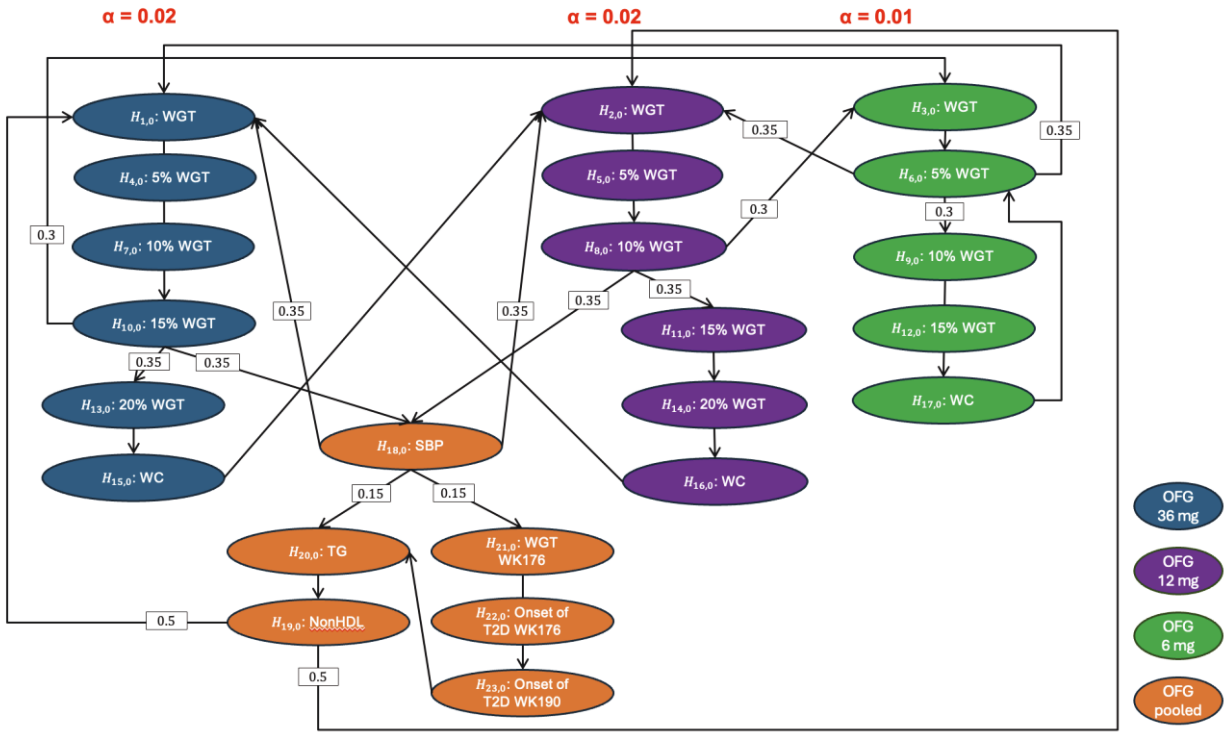
- $H_{4,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction from baseline at Week 72.
- $H_{5,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction from baseline at Week 72.
- $H_{6,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 5% or more body weight reduction from baseline at Week 72.
- $H_{7,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 10% or more body weight reduction from baseline at Week 72.
- $H_{8,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 10% or more body weight reduction from baseline at Week 72.
- $H_{9,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 10% or more body weight reduction from baseline at Week 72.
- $H_{10,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 15% or more body weight reduction from baseline at Week 72.
- $H_{11,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 15% or more body weight reduction from baseline at Week 72.
- $H_{12,0}$: No difference in 6 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 15% or more body weight reduction from baseline at Week 72.
- $H_{13,0}$: No difference in 36 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 20% or more body weight reduction from baseline at Week 72.

- $H_{14,0}$: No difference in 12 mg orforglipron compared to placebo with respect to the percentage of participants who achieve 20% or more body weight reduction from baseline at Week 72.
- $H_{15,0}$: No difference in 36 mg orforglipron compared to placebo with respect to mean change from baseline in waist circumference at Week 72.
- $H_{16,0}$: No difference in 12 mg orforglipron compared to placebo with respect to mean change from baseline in waist circumference at Week 72.
- $H_{17,0}$: No difference in 6 mg orforglipron compared to placebo with respect to mean change from baseline in waist circumference at Week 72.
- $H_{18,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to mean change from baseline in systolic blood pressure (SBP) at Week 72.
- $H_{19,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to mean percent change from baseline in non-high-density lipoprotein (HDL) cholesterol at Week 72.
- $H_{20,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to mean percent change from baseline in triglycerides at Week 72.
- $H_{21,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to mean percent change from baseline in body weight at Week 176 in participants with prediabetes at randomization.
- $H_{22,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to time to onset of T2D up to Week 176 in participants with prediabetes at randomization.
- $H_{23,0}$: No difference in 6 mg, 12 mg, and 36 mg orforglipron pooled compared to placebo with respect to time to onset of T2D up to Week 190 in participants with prediabetes at randomization.

2.1. Multiplicity Adjustment

Multiplicity adjusted analyses will be performed on the primary and key secondary objectives to control the overall family-wise Type 1 error rate at a 2-sided alpha level of 0.05. The graphical multiple testing procedure described in Bretz and colleagues (2009, 2011) will be used. This approach is a closed testing procedure, hence, it strongly controls the family-wise Type 1 error rate across all hypotheses (Alosh et al. 2014).

Figure GZGP.2.1 illustrates the final graphical testing scheme, including testing order, interrelationships, Type 1 error allocation for the primary and key secondary objectives, and the associated propagation. The treatment regimen estimand will be the primary estimand, with the efficacy estimand considered supportive. Since these estimands are intended for distinct purposes, no multiplicity adjustment will be made for conducting separate analyses on the same objectives. Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses outside the primary and key secondary objectives.



Abbreviations: OFG = orforglipron.

Figure GZGP.2.1. Graphical testing procedure.

3. Analysis Sets

For the purpose of analysis, the analysis populations are defined in Table GZGP.3.1 along with their attributes.

Table GZGP.3.1. Participant Analysis Sets

Participant Analysis Set	Description
Entered participants	**Definition:** All participants who sign informed consent[a]. **Purpose:** Used for providing summaries for screen failures and reasons associated with screen failures. Treatment Groups: None Inferential Comparisons: None
Randomized participants	**Definition:** All participants who are randomly assigned a study intervention. **Purpose:** Used for listings of disposition and treatment assignment summaries of disposition, demographics, historical illness, preexisting conditions, and analyses of efficacy and health outcomes. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO; OFG pooled vs. PBO.
Safety participants	**Definition:** All participants who are randomly assigned a study intervention and who take at least 1 dose of study intervention. **Purpose:** Used for all safety analyses. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO.
Randomized participants with prediabetes	**Definition:** All participants who are randomly assigned a study intervention and who have prediabetes at randomization[b]. **Purpose:** Used for listings of disposition and treatment assignment summaries of disposition, demographics, historical illness, preexisting conditions, and analyses of efficacy and health outcomes. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO; OFG pooled vs. PBO.
Safety participants with prediabetes	**Definition:** All participants who are randomly assigned a study intervention and take at least 1 dose of study intervention and have prediabetes at randomization[b]. **Purpose:** Used for all safety analyses. **Treatment Groups:** OFG 6 mg, OFG 12 mg, OFG 36 mg, PBO **Inferential Comparisons:** OFG 6 mg vs. PBO; OFG 12 mg vs. PBO; OFG 36 mg vs. PBO.

Abbreviations: IWRS = interactive web response system; OFG = orforglipron; PBO = placebo.

[a] Refers to the informed consent for the study.

[b] Determined by confirmed IWRS data .

The following data points sets are defined in Table GZGP.3.2 for all parameters:

Table GZGP.3.2. Data Points Sets

Data Points Sets	Description
Treatment regimen estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the last visit within the treatment period, regardless of study intervention discontinuation or initiation of prohibited weight management treatments. Baseline is defined in Section 4.1.1.
Efficacy estimand data points set	All data points obtained during the treatment period defined as at or after baseline and up to the earliest date of discontinuation of study intervention or initiation of prohibited weight management treatments. Baseline is defined in Section 4.1.1. and the prohibited weight management treatments are specified in Section 6.3 and Table GZGP.6.5.
Safety data points set	All data points obtained during the treatment period and follow-up period defined as at or after baseline and up to the date of study withdrawal or study completion including the follow-up period, regardless of intercurrent events. Baseline is defined in Section 4.1.1.

Note: at the time of the primary database lock, the treatment period is defined from Week 0 visit to Week 72 visit; for safety analyses, the additional 2-year treatment period and the corresponding follow-up period for prediabetes participants will not be included.

For the analysis of change from baseline to Week 190 for selected efficacy endpoints in randomized participants with prediabetes, the analyses will be conducted using a modified efficacy estimand data points set and the safety data points set. The modified efficacy estimand data points set includes all data points obtained during the 176-week treatment period and the follow-up visit at Week 190 defined as at or after baseline and up to the earliest date of discontinuation of study intervention or initiation of prohibited weight management treatments.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in SAP GZGP or the clinical study report (CSR). Additional exploratory data analyses may be conducted as deemed appropriate.

Details about the analyses regarding demographic and baseline characteristics, historical illnesses and preexisting conditions, treatment compliance, concomitant medications, and important protocol deviations can be found in Appendices 1 through 5 (Section 6.1 through Section 6.5), respectively.

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). Not all analyses described in SAP GZGP will necessarily be included in the CSRs. Any analyses described in this SAP and not provided in the CSR will be available upon request. Not all displays will necessarily be created as a "static" display. Some may be incorporated into interactive display tools instead of, or in addition to, a static display.

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

The change from baseline will be calculated as the value of interest at the visit minus the baseline value. In general, percent change from baseline will be calculated as the value of change from baseline divided by the baseline value in 100% scale. For some specific predefined parameters, percent change from baseline might be calculated using a log transformation. If the baseline value is missing for a particular variable, then the change from baseline and percent change from baseline will not be calculated when deriving the summary statistics.

For stratification factors at baseline, a strata variable is defined for statistical modeling to consist of 4 joint levels. The strata variable includes sex (female, male) and prediabetes status (yes, no). For analyses for participants with prediabetes at randomization, the strata variable will only include sex. Although country is also a stratification factor, geographic region (North America, Europe, Asia, or Central/South America) will be included in statistical modeling as a separate factor to improve model convergence.

Data may exist at visits where a variable was not scheduled to be collected, due to, for example, discontinuation visits. In these situations, data from the early discontinuation visit that does not correspond to the planned collection schedule will be excluded from the mixed model for repeated measures (MMRM), analysis of covariance (ANCOVA), or logistic regression analysis, unless otherwise specified (Andersen and Millen 2013).

Handling of missing data is addressed in Section 4.1.2 and Table GZGP.4.6. Section 3 provides definitions for the participant analysis set and data points set which includes definitions for censored data where appropriate. Handling spurious data is addressed in Section 6.5. Section 6.5 also addresses important protocol deviations. Protocol GZGP Section 10.1.7 addresses data quality assurance.

4.1.1. Definition of Baseline

Unless otherwise specified, the baseline for efficacy assessments is defined as the last available non-missing measurement prior to the first dose of study intervention; in most cases, this will be the measurement recorded at Week 0 (Visit 3). If there are no doses of study intervention administered, the baseline will be defined as the last available non-missing measurement on or prior to randomization. In cases where the measurement is taken on the same day (where the time is not collected or not reliable) as the first dose, this measurement will be used as the baseline value for data analysis. For patient-reported outcome measures data obtained at Visit 3, regardless of the timing relative to first dose, will serve as the baseline.

For safety assessments, the definition of baseline and postbaseline are specified in Table GZGP.4.1.

Table GZGP.4.1. Baseline and Postbaseline Definitions for Safety Analyses

Analysis Type	Baseline Period	Postbaseline Period
TEAEs	Starts from informed consent date and ends prior to the first dose (typically at Week 0).	Starts after the first dose of study intervention[a] and ends at the end of the follow-up period[b], or the date of study withdrawal.
TE abnormal laboratory values, vital signs, and ECGs	Starts from informed consent date and ends prior to the first dose (typically at Week 0). All scheduled and unscheduled measurements will be included.	Starts after the first dose and ends at the end of the follow-up period[b] or the date of study withdrawal. All scheduled and unscheduled measurements will be included.
Change from last baseline to each postbaseline week and to last postbaseline for laboratory values, vital signs, and ECGs	When "last baseline" is used, starts from informed consent date and ends prior to the first dose (typically at Week 0).	Starts after the first dose and ends at the end of the follow-up period[b] or the date of study withdrawal. Only scheduled visits and early termination visits that fall on the scheduled visits will be included.

Abbreviations: ECG = electrocardiogram; TE = treatment-emergent; TEAE = treatment-emergent adverse events.

[a] For events occurring on the day of first dose, information collected from the case report form will be used to determine whether the event was pre- versus posttreatment, if available. If the relevant information is not available, then the events will be counted as postbaseline.

[b] At Week 74 visit for participants without prediabetes or Week 72 visit if continuing in the trial for prediabetes participants at the primary database lock; and at Week 190 for prediabetes participants at final database lock.

4.1.2. Analysis Methods

The analysis methods are consistent with the desired estimands and the FDA guidance on "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance Document" (FDA 2023).

4.1.2.1. Analysis Methods for Treatment Regimen Estimand

4.1.2.1.1. Method for Continuous Variables

The ANCOVA model will be used to analyze continuous measurements at Week 72 and Week 176 and will be guided by the treatment regimen estimand. The model will include

- treatment group as a factor variable
- region as a factor variable
- strata as a factor variable
- baseline value (of the dependent variable)
- interaction between strata variable and treatment group, and
- interaction between baseline value and treatment group.

The estimated treatment group effect and comparison between 6 mg orforglipron, 12 mg orforglipron, and 36 mg orforglipron versus placebo will be reported together, with variability estimated using the robust inference (Ye et al. 2022; FDA 2023). The associated 2-sided 95% confidence interval and corresponding p-values will also be reported. If the model fails to

converge, all the interaction terms will be removed before the model fitting. The addition of interaction terms is not intended to estimate the heterogeneity effect but to provide robustness and efficiency for the estimate of treatment comparisons on the unconditional effect. The final inference will be derived using Rubin's Rule which combines estimates from multiple imputed datasets. The imputation procedure for creating a single imputed dataset is detailed in Table GZGP.4.2.

For some variables, both the baseline and the postbaseline values will be log transformed before fitting the ANCOVA. The treatment group estimates will be the percent change from baseline; the treatment contrasts will be the relative change in orforglipron 6 mg, orforglipron 12 mg, or orforglipron 36 mg compared to the placebo (%). In these cases, least squares (LS) means and 95% confidence intervals for each treatment group and treatment difference will be back-transformed and presented as mean percent change from baseline and relative treatment difference to placebo in percent change.

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

4.1.2.1.2. Primary Multiple Imputation (PMI) Strategy

Participants who discontinue study intervention (that is, discontinue study treatment) will be encouraged to continue in the study for the treatment period and follow-up period (note, the case report forms [CRFs] use "phase", which is interchangeable with the "period" in SAP). In this section, study discontinuation refers to treatment phase discontinuation captured in CRF.

If there are occurrences of missing data despite the best precautions, missing data should be imputed in a fashion consistent with what the values would likely have been had they been collected. In general, 5 scenarios are considered regarding the treatment journey of a trial participant relative to primary time point (Week 72 visit) shown in Figure GZGP.4.1.The same scenarios are considered for other relevant time points (e.g. Week 176).

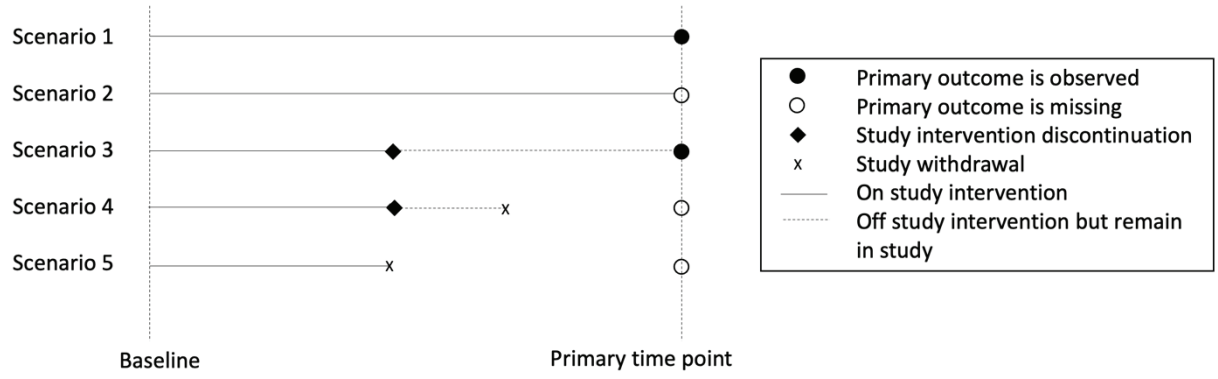


Figure GZGP.4.1. Treatment journey of a trial participant.

Scenario 5 covers the situations where a participant discontinues from the study intervention and then comes back for a study discontinuation visit for the same discontinuation reason on or prior to a scheduled visit. Additional details on scenario 5 are provided in Table GZGP.4.3.

In principle, missing data due to permanent discontinuation of study intervention (Scenario 4 and Scenario 5A) will be imputed by treatment group using retrieved dropouts (MI-RD), namely using multiple imputation based on data retrieved from participants who permanently discontinued the study intervention but continued in the study with non-missing measurements from the same treatment group. For scenarios 4 and 5A, if there are not enough retrieved dropouts to provide a reliable imputation model (that is, the model implemented does not converge), an alternative multiple imputation method (placebo-washout method) (Wang 2023) will be used. Any missing values at baseline will be imputed in the same model when imputing the endpoint.

Table GZGP.4.2. Imputation Procedure

Scenario (Study Intervention/Treatment Period Discontinuation; Missingness Status at Endpoint)	Methods to Handle Missing Values at Endpoint
1. No study intervention discontinuation; no missing value	N/A
2. No study intervention discontinuation; with missing value	Use observed data, including baseline and all postbaseline visits, across all time points from Scenario 1 and 2 to impute under the MAR assumption by treatment group. There should be very few such cases in a clinical trial. No additional covariates will be included in the imputation model.
3. Study intervention discontinuation; no missing value	N/A
4. Study intervention discontinuation with treatment period discontinuation at a subsequent visit or completion of treatment period; with missing value	Missing values at the endpoint visit will be imputed through MCMC (predictive mean matching method), using baseline and the endpoint visit data from Scenario 3 (MI-RD) by treatment group. No additional covariates will be included in the imputation model.
5. Study discontinuation resulting in study intervention discontinuation; with missing value	5A: If the study discontinuation is possibly related to study intervention (see treatment period discontinuation reasons classified as Category 5A in Table GZGP.4.3), missing values at the endpoint visit will be imputed in the same way as Scenario 4. 5B: If the study discontinuation is clearly due to administrative reasons not related to study intervention (see treatment period discontinuation reasons classified as Category 5B in Table GZGP.4.3), missing values at the endpoint visit will be imputed in the same way as Scenario 2.

Abbreviations: MAR = missing at random; MCMC = Markov chain Monte Carlo; MI = multiple imputation; RD = retrieved dropouts; N/A = not applicable.

Table GZGP.4.3. Categorization of Treatment Period Discontinuation Reasons

Disposition Reason	Associated Sub-Categories	Category
Adverse Event		5A
Death		5A
Protocol Deviation		5A
Pregnancy		5A
Non-Compliance With Study Drug		5A
Lack of Efficacy	Investigator	5A
	Study subject	5A
Withdrawal by Subject	Concern about study procedures/perceived risks	5A
	Health insurance changes	5A
	Scheduling conflicts	5A
	Subject is moving or has moved	5A
	Personal issue unrelated to trial	5A
	Due to epidemic/pandemic	5B
	Other	5A
Physician Decision	Due to epidemic/pandemic	5B
	Other	5A
Study Terminated by Sponsor		5B
Site Terminated by Sponsor		5B
Study Terminated by IRB/ERB		5B
Lost to Follow-up		5A
Other		5A

Abbreviations: ERB = ethical review board; IRB = institutional review board.

Note: For participants that discontinued the study due to emergency unblinding, the study disposition will be classified based upon the reason leading to emergency unblinding (for example, Adverse Event [5A]).

4.1.2.1.3. Modified Multiple Imputation Strategy

A modified multiple imputation (mMI) analysis is planned as a sensitivity analysis to assess the robustness of the primary efficacy results. An additional analysis will be conducted and data observed from Scenario 3 will be used to impute Scenario 5B.

4.1.2.1.4. Multiple Imputation Based Tipping-Point Analysis

A multiple imputation-based tipping-point (MI-TP) analysis is planned as a sensitivity analysis to explore how different patterns of body weight change post treatment period discontinuation by different treatment groups could impact the treatment comparisons.

To start with, missing data are imputed according to the primary imputation strategy (PMI) as shown in Section 4.1.2.1.2. A penalty is then added to those imputed values at the Week 72 visit. The MI-TP analysis varies the magnitude of the penalties added for both treatment groups under comparison and evaluates the impact these would have on the Study GZGP conclusion. A 2-dimensional space of penalties will be assessed for orforglipron 6 mg, orforglipron 12 mg, or orforglipron 36 mg and placebo. An MI-TP analysis aims to evaluate the robustness of the superiority claim to the assumptions of using the observed data to impute the missing body weight in all treatment groups.

4.1.2.1.5. Method for Binary Variables

The binary outcomes will be analyzed with the following procedure:

- Using the same imputation strategy – PMI as in Section 4.1.2.1.2 on the underlying continuous endpoint that determines the binary value. For example, the continuous body weight values at Week 72 visit will be imputed first.
- Transform the observed and imputed continuous value to the binary value, for example, convert the continuous body weight values at Week 72 visit to whether the corresponding percent change from baseline meets a certain threshold (Ma et al. 2022).
- Fit a logistic regression model to the data with the following terms:
    - treatment group as a factor variable
    - region as a factor variable
    - strata as a factor variable
    - baseline value (of the dependent variable)
    - interaction between treatment group and strata, and
    - interaction between treatment group and baseline value.
- Provide estimates and inferences of unconditional treatment effects defined by the risk difference and/or relative risk based on the delta-method using the formula provided (Ye et al. 2023).
- Derive the final inference using Rubin's rule which combines estimates from multiple imputed datasets. The imputation procedure for creating a single imputed dataset is detailed in Table GZGP.4.2.

The statistical testing for risk difference will be used as the primary approach for treatment comparisons.

4.1.2.1.6. Analysis Methods for Time to Event Variables

For time-to-event analysis, the log-rank test stratified on the strata variable will be conducted for the comparison between treatment groups. A Cox proportional hazards model stratified by the strata variable with treatment group as a factor, will be used as the primary analysis to estimate the HR between the treatment groups and the corresponding confidence interval.

For time-to-event analyses, participants who have not experienced the event of interest will be censored at the participant's end of follow-up. The censoring date will be the earliest of

- date of death
- date of the end of the analysis period/withdrawal, and
- the last confirmed visit date before participant was lost-to-follow-up.

The Kaplan-Meier method will be used to estimate the cumulative event curve over time. Counts and proportions of participants who experience an event will be calculated by treatment group.

For time-to-event analysis, if a participant experiences the event of interest, then time-to-event for that specific event of interest will be the number of days between the date of randomization and the onset date of the event plus 1 day. If a participant does not experience the event, then time-to-event for that specific event of interest will be the number of days between the date of randomization and the date of the participant's end of follow-up plus 1 day. If a participant

experiences multiple events the date of the first event will be used, unless otherwise specified. In rare cases where randomization occurred prior to the Visit 3 date, the Visit 3 date will be used instead.

4.1.2.2. Analysis Methods for Efficacy Estimand

4.1.2.2.1. Method for Continuous Variables

A maximum likelihood-based MMRM will be used to analyze continuous measurements at each postbaseline visit guided by the efficacy estimand. Missing data should be minimized at the best precaution. The hypothetical strategy will be used to handle ICEs. For MMRM analysis, only data collected before the occurrence of any ICEs will be used in the MMRM analysis. Through the MMRM, the potential efficacy measures (after the ICEs) will be implicitly imputed as if participants did not have ICEs. Participant dose modification and interruption will not disqualify their measurements from being included into the model.

The MMRM model will include the following terms:

- treatment group as a factor variable
- visit as a factor variable
- region as a factor variable
- strata as a factor variable
- baseline value (of the dependent variable)
- interaction between treatment group, strata variable and visit, and
- interaction between treatment group, baseline value and visit.

The estimated treatment group effect and comparison between 6 mg orforglipron, 12 mg orforglipron, or 36 mg orforglipron and placebo at the scheduled visits will be reported together with the variability estimated using the robust inference (Wang and Du 2023). The sandwich estimator (Diggle et al. 1994) for the variance-covariance matrix will be used. The addition of 3-way interaction terms is not intended to estimate the heterogeneity effect but to provide robustness and efficiency for the estimate of treatment comparisons on the unconditional effect. The associated 2-sided 95% confidence interval and corresponding p-values will also be reported. An unstructured covariance matrix by each treatment group will be used to model the within-participant errors, assuming heteroscedasticity and the measurements for different participants are independent. If the model fails to converge, the model will be simplified to include the following terms, removing 3-way interactions and the heteroscedasticity assumption:

- treatment group as a factor variable
- visit as a factor variable
- region as a factor variable
- strata as a factor variable
- baseline value (of the dependent variable)
- interaction between treatment group and visit
- interaction between strata variable and visit, and
- interaction between baseline value and visit.

If the model still fails to converge, the following covariance structures will be tested in order for the simplified model:

- heterogeneous Toeplitz
- heterogeneous autoregressive
- heterogeneous compound symmetry
- homogeneous Toeplitz
- homogeneous autoregressive, and
- homogeneous compound symmetry.

The first covariance structure that converges will be used.

For some variables, both the postbaseline response variables and baseline variable will be log-transformed before fitting the MMRM. The treatment group estimates will be the percent change from baseline while the treatment contrasts will be the relative change in orforglipron 6 mg, orforglipron 12 mg, or orforglipron 36 mg compared to the placebo (%) over the scheduled visits.

If the data does not warrant the use of an MMRM model, then an ANCOVA model will be conducted, as described in Section 4.1.2.1.1. Missing values at the visit of interest will be imputed through multiple imputation using all non-missing data (excluding data collected after ICEs) from the same treatment group under the missing at random (MAR) assumption.

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

4.1.2.2.2. Method for Binary Variables

The binary outcomes will be analysed with the following procedure:

- Impute the missing continuous-valued measurements at the scheduled visits using the randomized participants with efficacy estimand data points set assuming MAR with multiple imputation.
- At the visit of interest, transform the observed and imputed continuous value to the binary value, for example, convert the continuous body weight value at a visit to whether the corresponding percent change from baseline meets a certain threshold.
- Fit a logistic regression model to the transformed binary data with the following terms:
    - treatment group as a factor variable-
    - region as a factor variable
    - strata as a factor variable
    - continuous baseline value (of the dependent variable)
    - interaction between treatment group and strata, and
    - interaction between treatment group and baseline value.
- Provide for the visit of interest the estimates and inferences of unconditional treatment effects defined by the risk difference and/or risk ratio based on the delta-method using the formula provided (Ye et al. 2023).

- Derive the final inference using Rubin's rule by combining estimates from multiple imputed datasets.

The statistical testing for risk difference will be considered the primary method for treatment comparisons.

4.1.2.2.3. Analysis Methods for Time to Event Variables

Similar methodology will be used as described in Section 4.1.2.1.6, with the exception of the censoring rules.

For time-to-event analyses, participants who have not experienced the event of interest will be censored at the participant's end of follow-up. The censoring date will be the earliest of

- date of death
- date of initiation of prohibited weight management treatments
- date of discontinuation of study intervention, and
- the last confirmed visit date before participant was lost-to-follow-up.

4.1.2.3. Analysis Methods for Safety

In general, safety assessments will be guided by an estimand comparing safety of orforglipron doses with placebo regardless of intercurrent events. Thus, safety analyses will be conducted using the safety participants based on data observed during treatment period plus the posttreatment follow-up period, that is, the period from first dose of treatment to the end of the follow-up visit or the date of study withdrawal. It is noted that for some safety endpoints, such as study intervention discontinuation due to adverse event (AE) and so forth, the analyses will be based on data observed during treatment period only.

Fisher's exact test will be used for treatment comparisons of percentages. The risk differences and 95% confidence intervals will be provided.

For selected continuous safety measures, unless otherwise specified, treatment group differences of mean change or percent mean change from baseline at all scheduled visits will be assessed via an MMRM using maximum likelihood, which will include the following terms:

- treatment group as a factor variable
- visit as a factor variable
- baseline value (of the dependent variable)
- interaction between visit and treatment group

If the data does not warrant the MMRM model, then an ANCOVA model with treatment group as a fixed effect and the continuous baseline value as a covariate will be used.

No explicit imputation will be conducted for safety measures. Some parameters (such as urinary albumin to creatinine ratio [UACR], p-amylase, and lipase) may be log-transformed before fitting the MMRM as specified in Section 4.1.2.1.1. In these cases, least squares (LS) means and 95% confidence intervals for each treatment group and treatment difference will be back-transformed and presented as mean percent change from baseline and as relative treatment difference of orforglipron treatment groups compared to placebo in percent change.

The Kaplan–Meier (KM) product limit method will be used to estimate the cumulative event-free survival rates over time for the time-to-event analyses. An unstratified Cox proportional hazards regression analysis will be used to compare hazard rates among treatments. An unstratified log-rank test will be used to calculate p-values. Time-to-event for the specific safety event of interest will be calculated from the date of first dose.

A logistic regression model will be used for hypoglycemia incidence for treatment comparisons for the overall postbaseline period. The model will include the fixed effects of treatment. Given the expected small number of hypoglycemic events in the study population, the rate of events will be analyzed using an empirical method for the overall postbaseline period.

Some safety analyses may be conducted after excluding data after the initiation of anti-hyperglycemic medications for those who develop T2D.

4.2. Participant Dispositions

The participant dispositions for the screening period, study intervention, the treatment period, and/or the follow-up period will be collected in the CRFs with the corresponding primary reason.

The study completion status is defined as

- at the end of the 72-week treatment period (when the primary endpoint is ascertained and the primary database is locked): for participants without prediabetes at randomization, completers will be considered as those who complete the treatment period (Week 72) and the follow-up visit; for participants with prediabetes at randomization, completers will be considered as those who complete Week 72; otherwise, participants will be considered as non-completers.
- at the end of the 176-week treatment period for participants with prediabetes at randomization, completers will be considered as those who complete the treatment period (Week 176) and the follow-up visit; otherwise, participants will be considered as non-completers.

The planned listings and summary tables for dispositions are provided in Table GZGP.4.4. No inferential analysis will be performed. Additionally, the planned listings and summary tables for dispositions will be provided for participants with prediabetes at randomization for the final lock.

Table GZGP.4.4. Listings and Summary Tables Related to Dispositions

Analysis	Population/Period
Summary of disposition (prior to randomization)	Entered participants
Patient allocation by region, country, and center/site	Entered participants
Summary of study and study intervention disposition	Randomized participants/TP + FP (TP for study treatment disposition)
Kaplan-Meier plot of time to study discontinuation	Randomized participants/TP + FP
Kaplan-Meier plot of time to study intervention discontinuation	Safety participants/TP
Kaplan-Meier plot of time to study intervention discontinuation due to AEs	Safety participants/TP
Listing of randomization	Randomized participants
Listing of randomized participants who were discontinued from the study intervention due to inadvertent enrollment	Randomized participants
Listing of study and study intervention disposition	Randomized participants

Abbreviations: FP = follow-up period; TP = treatment period.

Time-to-study intervention discontinuation will be calculated from the date of first dose. Time-to-study discontinuation will be calculated from the date of randomization.

4.3. Primary Endpoint/Estimand Analysis

4.3.1. Definition of endpoint(s)

The primary efficacy endpoint is percent change in body weight at 72 weeks. As specified in Table GZGP.4.5, the percent change in body weight is defined as:

(postbaseline body weight [kg] – baseline body weight [kg]) / baseline body weight [kg] × 100%.

4.3.2. Main analytical approach

The analytical approaches are specified in Section 4.1.2.1.1 and Section 4.1.2.1.2 (ANCOVA with PMI) and Section 4.1.2.2.1 (MMRM).

4.3.3. Sensitivity Analysis

The sensitivity analysis are specified in Section 4.1.2.1.3 (MI-TP).

4.3.4. Supplementary analyses

No supplementary analysis is planned for the primary endpoint.

4.4. Secondary Endpoints/Estimands Analysis

Table GZGP.4.5 includes the description and derivation of the primary and secondary endpoints for the efficacy measures and PROs.

Table GZGP.4.6 provides the detailed analyses including endpoint, estimand, data points set, method and imputation, population, and time point for the primary and secondary endpoints.

Table GZGP.4.5. Descriptions and Derivation of Efficacy Measures and Patient Reported Outcomes

Measure	Description	Variable	Derivation/Comment	Handling Missing Components
Body weight	Per Protocol GZGP Section 10.7: Body weight measurements should be done in a consistent manner using a calibrated electronic scale capable of measuring weight in kilograms to 1 decimal place. All weights for a given participant should be measured using the same scale, whenever possible, at approximately the same time in the morning after evacuation of bladder contents. Body weight will be measured in fasting state at all visits. If the participant is not fasting, the participant should be called in for a new visit within the visit window to have the fasting body weight measured.	Body weight (kg)	As measured	Single item, missing if missing
		Change from baseline in body weight (kg)	Calculated as: postbaseline body weight (kg) – baseline body weight (kg)	Missing if baseline or postbaseline value is missing
		Percent change from baseline in body weight (%)	Calculated as: (postbaseline body weight [kg] – baseline body weight [kg]) / baseline body weight [kg] × 100 (%)	Missing if baseline or postbaseline value is missing
		≥x% body weight reduction from baseline where, x = 5, 10, 15, 20	Response = yes if at least a x% reduction in body weight from baseline, that is, percent change from baseline in body weight ≤ -x%	Missing if baseline or postbaseline value is missing
BMI	BMI: Round to one decimal point. For example, a BMI of 29.9 kg/m^2 should not be rounded to 30.0 kg/m^2.	BMI (kg/m^2)	BMI will be calculated as: Weight (kg) / (height [m])^2	Missing if weight or height is missing
		Change from baseline in BMI (kg/m^2)	Calculated as: postbaseline BMI (kg/m^2) – baseline BMI (kg/m^2)	Missing if baseline or postbaseline value is missing
		BMI threshold value of <x kg/m^2 Where, x = 25, 27, 30, 35	Response = yes if the postbaseline BMI value <x kg/m^2	Missing if postbaseline value is missing
Waist circumference	Per Protocol GZGP Section 10.7: Waist circumference should be measured in the horizontal plane and at the midpoint between the lower margin of the last palpable rib and the top of the iliac crest.	Waist circumference (cm)	The average of 2 measures	If there is at least one measurement, take the average of all non-missing measurements; otherwise, set it to be missing

	Measurements should be taken at the end of a normal expiration using a non-stretchable measuring tape. The tape should lie flat against the skin without compressing the soft tissue. The waist circumference should be measured twice, rounded to the nearest 0.5 cm. The measuring tape should be removed between the 2 measurements. Both measurements will be recorded in the CRF. If the difference between the 2 measurements exceeds 1 cm, this set of measurements should be discarded and the 2 measurements repeated.	Change from baseline in waist circumference (cm)	Calculated as: postbaseline Waist circumference (cm) – baseline Waist circumference (cm)	Missing if baseline or postbaseline value is missing
Blood pressure	Per Protocol GZGP Section 10.7: Have the participant sit quietly for about 5 minutes before vital signs measurements are taken. For each parameter, take 3 measurements from the same arm, preferably the nondominant arm. Measure the recordings at least 1 minute apart. Blood pressure must be taken with an automated blood pressure instrument	SBP (mmHg)	The average of 3 measures	If there is at least one measurement, take the average of all non-missing measurements; otherwise, set it to be missing
		Change from baseline in SBP (mmHg)	Calculated as: postbaseline SBP (mmHg) – baseline SBP (mmHg)	Missing if baseline or postbaseline value is missing
		DBP (mmHg)	The average of 3 measures	If there is at least one measurement, take the average of all non-missing measurements; otherwise, set it to be missing
		Change from baseline in DBP (mmHg)	Calculated as: postbaseline DBP (mmHg) – baseline DBP (mmHg)	Missing if baseline or postbaseline value is missing
Lipid parameter	HDL-C, triglycerides, and total cholesterol will be assayed by Lilly designated laboratory. LDL-cholesterol, non-HDL-cholesterol, and VLDL-C will be calculated by Lilly designated laboratory.	Triglycerides	As provided. Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in triglycerides (%)	Calculated as: log(postbaseline triglycerides) – log (baseline triglycerides)	Missing if baseline or postbaseline value is missing

			then will transform back to percent change.	
		Total cholesterol	As provided. Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in total cholesterol (%)	Calculated as: log(postbaseline total cholesterol) – log(baseline total cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		Non-HDL-cholesterol	As provided. Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in non-HDL-cholesterol (%)	Calculated as: log(postbaseline non-HDL-cholesterol) – log(baseline non-HDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		LDL-cholesterol	As provided. If triglycerides are >400 mg/dL, the direct LDL will be directly measured Log transformation before the analysis.	Single item, missing if missing
		Percent change from baseline in LDL-cholesterol (%)	Calculated as: log(postbaseline LDL-cholesterol) – log(baseline LDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		HDL-cholesterol	As provided. Log transformation before the analysis	Single item, missing if missing

		Percent change from baseline in HDL-cholesterol (%)	Calculated as: log(postbaseline HDL-cholesterol) – log(baseline HDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
		VLDL-cholesterol	As provided. Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in VLDL-cholesterol (%)	Calculated as: log(postbaseline VLDL-cholesterol – log(baseline VLDL-cholesterol) then will transform back to percent change	Missing if baseline or postbaseline value is missing
Glycemic control	Fasting glucose will be assayed by Lilly-designated laboratory.	Fasting glucose (mg/dL and mmol/L)	As provided	Single item, missing if missing
		Change from baseline fasting glucose (mg/dL and mmol/L)	Calculated as: Postbaseline fasting glucose – Baseline fasting glucose	Missing if baseline or postbaseline value is missing
	HbA1c will be assayed by Lilly-designated laboratory.	HbA1c (% and mmol/mol)	As provided	Single item, missing if missing
		Change from baseline HbA1c (% and mmol/mol)	Calculated as: Postbaseline HbA1c – Baseline HbA1c	Missing if baseline or postbaseline value is missing
	Delayed progression to T2D is evaluated by time to onset of T2D (only for participants with prediabetes at randomization).	Time to onset of T2D	Calculated as: (date of onset of CEC adjudicated T2D event – date of randomization) + 1 day	Censored if no postbaseline event of T2D was observed
	Percentage of participants with prediabetes at randomization achieving normoglycemia.	Percentage of participants achieving normoglycemia	Response = Yes if FSG <100 mg/dL (obtained alone or at time = 0 during an OGTT), OGTT 2h SG <140 mg/dL, and HbA1c <5.7%.	Missing if postbaseline values of HbA1c, FSG and OGTT are missing

	2-hour OGTT will be performed per Protocol Section 10.7: • participants should attend visits in the fasting state, and • serum samples will be collected at 0, 30, 60, 90, and 120 minutes	Point x OGTT (mg/dL) x= 0, 30, 60, 90, 120 minutes	As provided	Single item, missing if missing
		Change from baseline in point x OGTT (mg/dL)	Calculated as: postbaseline point x OGTT (mg/dL) – baseline point x OGTT (mg/dL)	Missing if baseline or postbaseline value is missing
Insulin Sensitivity	Matsuda index will be derived from 2-hour OGTT results.	Matsuda index	Calculated as: $\frac{10000}{\sqrt{(FSG \times FPI) \times (\bar{G} \times \bar{I})}}$, where FSG is the fasting glucose, FPI is the fasting insulin, $\bar{G}$ denotes the mean glucose during OGTT, and $\bar{I}$ denotes the mean insulin during OGTT.	Missing if any values in the derivation formula is missing
		Change from baseline in Matsuda index	Calculated as: postbaseline Matsuda index – baseline Matsuda index	Missing if baseline or postbaseline value is missing
Fasting Insulin	Per Protocol GZGP Section 10.2: Fasting insulin will be assayed by Lilly-designated laboratory.	Fasting insulin	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in fasting insulin (%)	Calculated as: log(postbaseline fasting insulin) – log(baseline fasting insulin) then will transform back to percent change.	Missing if baseline or postbaseline value is missing
Renal Function	Per Protocol GZGP Section 10.2: UACR and eGFR will be assayed by Lilly-designated laboratory.	UACR	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in UACR (%)	Calculated as: log(postbaseline UACR) – log(baseline UACR) then will transform back to percent change.	Missing if baseline or postbaseline value is missing

		eGFR CKD-EPI Cystatin C (ml/min/1.73m^2)	As provided as: eGFR (CKD-EPI Cystatin C 2012)	Single item, missing if missing
		Change from baseline in eGFR CKD-EPI Cystatin C (ml/min/1.73m^2)	Calculated as: postbaseline eGFR CKD-EPI Cystatin C (ml/min/1.73m^2) – Baseline eGFR CKD-EPI Cystatin C (ml/min/1.73m^2)	Missing if baseline or postbaseline value is missing
		eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	As provided as: eGFR (CKD-EPI Creatinine 2021)	Single item, missing if missing
		Change from baseline in eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	Calculated as: postbaseline eGFR CKD-EPI Creatinine (ml/min/1.73m^2) – Baseline eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	Missing if baseline or postbaseline value is missing
Inflammatory biomarker	Per Protocol GZGP Section 10.2: hsCRP will be assayed by Lilly-designated laboratory.	hsCRP (mg/L)	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in hsCRP (%)	Calculated as: log(postbaseline hsCRP) – log(baseline hsCRP)	Missing if baseline or postbaseline value is missing
FFAs	Per Protocol GZGP Section 10.2: FFAs will be assayed by Lilly-designated laboratory.	FFAs (mmol/L)	As provided Log transformation before the analysis	Single item, missing if missing
		Percent change from baseline in FFAs (%)	Calculated as: log(postbaseline FFAs) – log(baseline FFAs)	Missing if baseline or postbaseline value is missing
HOMA	HOMA2 B estimates steady state beta cell function	HOMA2 B (insulin)	Calculated from the link[a] using fasting glucose and insulin laboratory values	Missing if fasting glucose or insulin is missing
		Percent change from baseline in HOMA2 B (insulin) (%)	Calculated as: log[postbaseline HOMA2 B (insulin)] – log[baseline HOMA2 B (insulin)]	Missing if baseline or postbaseline value is missing

		HOMA2 B (C-peptide)	Calculated from the link[a] using fasting glucose and C-peptide laboratory values	Missing if fasting glucose or C-peptide is missing
		Percent change from baseline in HOMA2 B (C-peptide) (%)	Calculated as: log [postbaseline HOMA2 B (C-peptide)] – log [baseline HOMA2 B (C-peptide)]	Missing if baseline or postbaseline value is missing
	HOMA2 IR estimates insulin resistance	HOMA2 IR (insulin)	Calculated from the link[a] using fasting glucose and insulin laboratory values	Missing if fasting glucose or insulin is missing
		Percent change from baseline in HOMA2 IR (insulin) (%)	Calculated as: log[postbaseline HOMA2 IR (insulin)] – log[baseline HOMA2 IR (insulin)]	Missing if baseline or postbaseline value is missing
		HOMA2 IR (C-peptide)	Calculated from the link[a] using fasting glucose and C-peptide laboratory values	Missing if fasting glucose or C-peptide is missing
		Percent change from baseline in HOMA2 IR (C-peptide) (%)	Calculated as: log[postbaseline HOMA2 IR (C-peptide)] – log[baseline HOMA2 IR (C-peptide)]	Missing if baseline or postbaseline value is missing
SF-36	SF-36v2 will assess health-related quality of life. The SF-36v2 Acute Form, 1-week recall version is a 36-item generic, participant-completed measure designed to assess the following 8 domains. • Physical functioning • Role-physical • Bodily pain • General health • Vitality • Social functioning • Role-emotional, and • Mental health.	SF-36 domain scores and SF-36 component scores	Per copyright owner, the Quality Metric Health Outcomes™ Scoring Software will be used to derive SF-36 domain and component scores. After data quality-controls, the SF-36 software will recalibrate the item-level responses for calculation of the domain and component scores. These raw scores will be transformed into the domain scores (T-scores) using the 1-week recall period. This entails exporting the patient data in a CSV or tab-	Missing data handling offered by SF-36 software will be used. "Maximum Data Recovery" will be selected for Missing Score Estimator in the software.

	The Physical Functioning domain assesses limitations due to health now. The Physical functioning domain does not have a recall period while the remaining domains assess functioning "in the past week." Each domain is scored individually and information from these 8 domains is further aggregated into 2 health component summary scores, a Physical Component Summary, and a Mental Component Summary. Items are answered on Likert scales of varying lengths (3-point, 5-point, or 6-point scales). Scoring of each domain and both summary scores are norm based and presented in the form of T-scores, with a mean of 50 and SD of 10. Higher scores indicate better levels of function and/or better health[b].		delimited file for import, generation of the SF-36 scores and reports, and export of the calculated scores in a CSV or tab-delimited file.	
		SF-36 change from baseline for domains and component scores (PCS and MCS)	Calculated as: postbaseline SF-36 score – baseline SF-36 score	Missing if baseline or postbaseline value is missing
PGI-S	The PGI-S Physical Function Weight Scale is designed to assess the participants' overall perception of their condition. This is a single global item that asks participants to rate how their weight limited their ability to perform physical activities in the past 7 days.	PGI-S rating	As assessed. The responses are on a 5-point scale ranging from "not at all limited" to "extremely limited."	Single item, missing if missing
PGI-C	The PGI-C Physical Function Weight Scale is designed to assess the participants' overall perception of the efficacy of treatment. This is a single global item that asks participants to rate the overall change in their ability to perform physical activities due to their weight since starting the study intervention.	PGI-C rating	As assessed. The responses are based on a 5-point scale ranging from "much better" to "much worse."	Single item, missing if missing

IWQOL-Lite-CT	IWQOL-Lite-CT[c] is a 20-item, obesity-specific, patient-reported outcome instrument developed for use in weight management clinical studies. The IWQOL-Lite-CT assesses 2 primary domains of obesity-and health-related quality of life: Physical (7 items) and Psychosocial (13 items), and a 5-item subset of the Physical domain, the Physical Function composite. Items in the Physical Function composite describe physical impacts related to general and specific physical activities. All items are rated on either a 5-point frequency ("never" to "always") scale or a 5-point truth ("not at all true" to "completely true") scale. The 2 domain scores, 1 composite score and total score range from 0 to 100 with higher scores indicating greater functioning, derived as below: Raw scores for each subscale are computed if a minimum of 50% of the items for that subscale are non-missing, and for the IWQOL-Lite-CT total score if a minimum of 75% of all items are non-missing. If the minimum number of items is answered for a composite, then • compute the average of the valid non-missing responses corresponding to the items in the total or each domain/composite will be calculated (1 = "never" or "not at all true" and 5 = "always" or "completely true"). • The score will be then calculated by transforming the raw score to the 0 (worst) to 100 (best) metric using the following	Physical domain score and psychosocial domain score	Physical domain includes Items 1-5, 16, 17; Psychosocial domain includes Items 6-8, 9-15, 18, 19, 20	The minimum requirements for non-missing responses: Psychosocial: 7 of 13 items, Physical: 4 of 7 items, otherwise considered missing
		Change from baseline in physical domain score and psychosocial domain score	Calculated as: Postbaseline score – baseline score	Missing if baseline or postbaseline value is missing
		Physical function composite score	Include Items 1-3, 16, 17.	The minimum requirements for non-missing responses: 3 of 5 items, otherwise considered missing
		Change from baseline in physical function composite score	Calculated as: postbaseline score – baseline score	Missing if baseline or postbaseline value is missing
		Total score	Items 1-20	The minimum requirements for non-missing responses: 15 of 20 items, otherwise considered missing
		Change from baseline in total score	Calculated as: Postbaseline score – baseline score	Missing if baseline or postbaseline value is missing

	formula for every participant at each time point: $100(S_{max} - C_{avg})/(S_{max} - S_{min})$ ○ C_{avg} is the raw average score of all non-missing item responses in the composite; this average must be a number between 1 and 5, inclusive. ○ S_{max} is the maximum possible raw score value (that is, 5) ○ S_{min} is the minimum possible raw score value (that is, 1) ○ Inserting the maximum and minimum possible score values, the formula is reduced to: $100(5 - C_{avg}) / 4$			
EQ-5D-5L and VAS	The EQ-5D-5L[d] is a standardized, 5-item, self-administered instrument for use as a measure of health outcome. It provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care as well as population health surveys. The EQ-5D-5L assesses 5 dimensions of health: Item 1: mobility Item 2: self-care Item 3: usual activities Item 4: pain/discomfort Item 5: anxiety/depression Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The VAS records the respondent's self-rated health on a vertical VAS where the endpoints	EQ-5D-5L item scores	Five health profile dimensions, each dimension has 5 levels: 1 = no problems 2 = slight problems 3 = moderate problems 4 = severe problems 5 = extreme problems It should be noted that the numerals 1-5 have no arithmetic properties and should not be used as a primary score.	Each dimension is a single item, missing if missing. Note: Missing value can be coded as 9.
		EQ-5D-5L UK population-based utility index	EQ-5D-5L health states are converted into a single index "utility" score using a scoring algorithm based on public preferences. Uses the concatenation of the value of each EQ-5D-5L dimension score	If any of the items is missing or coded as 9, the index score is missing

	are labeled as “best imaginable health state” (100) and “worst imaginable health state” (0).		in the order: item 1; item 2; item 3; item 4; item 5. Derive EQ-5D-5L UK population-based index score according to the link[b] by using the UK algorithm to produce a patient-level index score between -0.59 and 1.0 (continuous variable)[e].	
		Change from baseline in EQ-5D-5L utility index	Calculated as: Postbaseline utility index score – baseline utility index score	Missing if baseline or postbaseline value is missing
		EQ-5D VAS	Range from 0 = “worst imaginable health state” to 100 = “best imaginable health state” Note: Higher value indicates better health state.	Single item, missing if missing.
		Change from baseline in EQ-5D-5L VAS	Calculated as: postbaseline VAS score – baseline VAS score	Missing if baseline or postbaseline value is missing
PROMIS SD	The PROMIS Short Form v1.0 Sleep Disturbance 8b assesses self-reported perceptions of sleep quality, sleep depth, and restoration associated with sleep, including perceived difficulties and concerns with getting to sleep or staying asleep, as well as perceptions of the adequacy of and satisfaction with sleep. The PROMIS Short Form v1.0 Sleep Disturbance 8b has a recall period of 7 days and each of its 8 items are rated on a 5-point scale ranging from “not at all” to “very much”, “never” to “always,” or “very poor” to “very good.” Higher T-scores indicate more sleep disturbance.	PROMIS SD T-scores	T-scores are generated using response pattern scoring with a mean of 50 and a SD of 10[f].	Missing if all of the items are missing
		Change from baseline in PROMIS SD T-scores	Calculated as: postbaseline score – baseline score	Missing if baseline or postbaseline value is missing

Appetite VAS	The aim of the appetite VAS is to determine the effects of study intervention on appetite sensations and desire for specific foods. Participants will be asked to rate their feelings of hunger, satiety, fullness, prospective food consumption, and desire for specific foods by making a vertical mark on a 10-cm (100 mm) line. The ratings will include the following 8 questions[g]: • How hungry do you feel right now? • How satisfied do you feel right now? • How full do you feel right now? • How much food do you think you could eat right now? • Would you like to eat something sweet? • Would you like to eat something salty? • Would you like to eat something savory? • Would you like to eat something fatty?	Appetite VAS item scores	For questions 1-4: range from 0 mm = "not at all" to 100 mm = "extremely" For questions 5-8: range from 0 mm = " No, not all " to 100 mm = "Yes, very much"	Single item, missing if missing
		Appetite VAS overall score	Calculated as the mean of the following 4 fasting appetite VAS item scores: • How hungry do you feel right now? • How satisfied do you feel right now? • How full do you feel right now? • How much food do you think you could eat right now? (van Can et al. 2014) A higher value means feeling hungry or strong desire to eat.	Missing if any single item is missing
		Change from baseline in appetite VAS item and overall scores	Calculated as: postbaseline VAS score – baseline VAS score	Missing if baseline or postbaseline value is missing
		Change from Week 72 in appetite VAS item and overall scores	Calculated as: postbaseline VAS score – Week 72 VAS score	Missing if Week 72 or postbaseline value is missing
PFS	The PFS[h] is a 15-item, self-administered instrument that assesses the psychological impact of living in an environment with an abundance of palatable foods. It measures the	PFS item scores	5-point Likert scale ranging from 1 (do not agree at all) to 5 (strongly agree) Higher scores indicate a higher psychological impact of food.	Single item, missing if missing

	appetite for food based on 3 levels of food proximity: • food available • food present, and • food tasted.	PFS domain scores	Calculated as the mean of the item scores within each of the individual 3 domains	Missing if any single item is missing
		PFS overall score	Calculated as the mean of all 15 items	Missing if any single item is missing
		Change from baseline in PFS domain and overall scores	Calculated as: postbaseline PFS score – Baseline PFS score	Missing if baseline or postbaseline value is missing
		Change from Week 72 in PFS domain and overall scores	Calculated as: postbaseline PFS score – Week 72 PFS score	Missing if baseline or postbaseline value is missing

Abbreviations: BMI = body mass index; CRF = case report form; CSV = comma separated values; DBP = diastolic blood pressure; eGFR = estimated glomerular filtration rate; FFA = free fatty acids; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; HOMA = homeostasis model assessment; IWQOL-Lite-CT = Impact of Weight on Quality of Life-Lite Clinical Trials Version; HOMA = homeostasis model assessment; HOMA2 B = homeostasis model assessment 2 B; HOMA2 IR = homeostasis model assessment 2 IR; hsCRP = high-sensitivity C-reactive protein; LDL = low-density lipoprotein; MCS = Mental Component Score; OGTT = oral glucose tolerance test; PGI--C = patient global impression of change; PCS = Physical Component Score; PFS = Power of Food Scale; PGI-S = patient global impression of severity; PROMIS SD = Patient-Reported Outcomes Measurement Information System Sleep Disturbance; SBP = systolic blood pressure; SF-36 = Short Form 36-item Health Survey; SF-36v2 = Short Form 36 Health Survey Version 2; T2D = type 2 diabetes; UACR = urine albumin-creatinine ratio; VAS = visual analog scale; VLDL = very low-density lipoprotein.

[a] Derive HOMA2 B and HOMA2 IR using the calculator at https://www.rdm.ox.ac.uk/about/our-clinical-facilities-and-mrc-units/DTU/software/homa

[b] Maruish 2011.

[c] Kolotkin et al. 2017, 2019.

[d] EuroQol Research Foundation 2019.

[e] Derive EQ-5D-5L UK population-based index score using the calculator at https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/crosswalk-index-value-calculator/

[f] Scoring algorithm obtained from the developer, Northwestern University (Chapman 2022).

[g] Flint et al. 2000.

[h] Cappelleri et al. 2009; Lowe et al. 2009.

Table GZGP.4.6. Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Estimand Data Points Set / Population[a]	Analysis Method	Time Point[b]	Analysis Type
Body weight	Percent change from baseline in body weight (%)	Treatment Regimen	ANCOVA with PMI	Week 72 visit	Primary endpoint, main analysis
		Efficacy	MMRM	Week 72 visit	Primary endpoint, main analysis
		Treatment Regimen	ANCOVA with MI-TP	Week 72 visit	Primary endpoint, sensitivity analysis
		Treatment Regimen	Cumulative distribution function (CDF) plot with PMI	Week 72 visit	Exploratory analysis
		Efficacy	CDF plot with MI	Week 72 visit	Exploratory analysis
		Treatment Regimen	Waterfall Plot with observed data	Week 72 visit	Exploratory analysis
		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI	Week 176 visit	Key secondary, main analysis
		Efficacy / Randomized participants with prediabetes	MMRM	Week 176 visit	Key secondary, main analysis
		Modified efficacy/Randomized participants with prediabetes	MMRM	Week 190 visit	Exploratory analysis
		Safety /Randomized participants with prediabetes	MMRM	Week 190 visit	Exploratory analysis
	Percentage of participants with ≥x% body weight reduction from baseline, where x = 5, 10, 15, 20	Treatment Regimen	Logistic regression with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	Logistic regression with MI	Week 72 visit	Key secondary, main analysis

		Treatment Regimen / Randomized participants with prediabetes	Logistic regression with PMI	Week 176	Additional secondary (x = 5); Exploratory analysis (x = 10, 15, 20)
		Efficacy / Randomized participants with prediabetes	Logistic regression with MI	Week 176	Additional secondary (x = 5); Exploratory analysis (x = 10, 15, 20)
	Change from baseline in body weight (kg)	Efficacy	MMRM	Week 72 visit	Additional secondary
		Treatment Regimen	ANCOVA with PMI	Week 72 visit	Additional secondary
		Efficacy / Randomized participants with prediabetes	MMRM	Week 176 visit	Exploratory analysis
		Modified efficacy/Randomized participants with prediabetes	MMRM	Week 190	Exploratory analysis
		Safety /Randomized participants with prediabetes	MMRM	Week 190 visit	Exploratory analysis
BMI	Change from baseline in BMI (kg/m^2)	Efficacy	MMRM	Week 72 visit	Additional secondary

		Treatment Regimen	ANCOVA with PMI	Week 72 visit	Additional secondary
		Efficacy / Randomized participants with prediabetes	MMRM	Week 176 visit	Exploratory analysis
	Percentage of participants with BMI threshold value of <x kg/m^2, where x = 25, 27, 30, 35	Efficacy	Logistic regression with MI	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	Logistic regression with MI	Week 176 visit	Exploratory analysis
		Treatment Regimen	Logistic regression with PMI	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	Logistic regression with PMI	Week 176 visit	Exploratory analysis
Waist circumference	Change from baseline in waist circumference (cm)	Treatment Regimen	ANCOVA with PMI	Week 72 visit	Key secondary, main analysis
		Efficacy	MMRM	Week 72 visit	Key secondary,

					main analysis
		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI	Week 176 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM	Week 176 visit	Exploratory analysis
		Modified efficacy/Randomized participants with prediabetes	MMRM	Week 190	Exploratory analysis
		Safety /Randomized participants with prediabetes	MMRM	Week 190 visit	Exploratory analysis
Blood pressure	Change from baseline in SBP (mm Hg) Change from baseline in DBP (mm Hg)	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analysis

		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
		Modified efficacy/Randomized participants with prediabetes	MMRM (pooled and individual OFG doses)	Week 190	Exploratory analysis
		Safety /Randomized participants with prediabetes	MMRM (pooled and individual OFG doses)	Week 190 visit	Exploratory analysis
Lipid parameter	Percent (%) change from baseline in • triglycerides • non-HDL-cholesterol • total cholesterol • LDL-cholesterol • HDL-cholesterol • VLDL-cholesterol	Treatment Regimen	ANCOVA with PMI[c] (log transformation; pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analyses
		Efficacy	MMRM (log transformation; pooled and individual OFG doses)	Week 72 visit	Key secondary, main analysis; exploratory analyses

		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI[c] (log transformation; pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (log transformation; pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
		Modified efficacy/Randomized participants with prediabetes	MMRM (log transformation; pooled and individual OFG doses)	Week 190	Exploratory analysis
		Safety /Randomized participants with prediabetes	MMRM (log transformation; pooled and individual OFG doses)	Week 190 visit	Exploratory analysis
Glycemic control	Change from baseline in fasting glucose (mg/dL and mmol/L)	Efficacy	MMRM	Week 72 visit	Additional secondary
		Treatment Regimen	ANCOVA with PMI	Week 72 visit	Additional secondary
		Efficacy / Randomized participants with prediabetes	MMRM	Week 176 visit	Additional secondary
		Modified efficacy/Randomized participants with prediabetes	MMRM	Week 190	Exploratory analysis

		Safety /Randomized participants with prediabetes	MMRM	Week 190 visit	Exploratory analysis
	Change from baseline in HbA1c (%)	Efficacy	MMRM	Week 72 visit	Additional secondary
		Treatment Regimen	ANCOVA with PMI	Week 72 visit	Additional secondary
		Efficacy / Randomized participants with prediabetes	MMRM	Week 176 visit	Additional secondary
		Modified efficacy/Randomized participants with prediabetes	MMRM	Week 190	Exploratory analysis
		Safety /Randomized participants with prediabetes	MMRM	Week 190 visit	Exploratory analysis
	Time to onset of T2D	Treatment Regimen / Randomized participants with prediabetes	Stratified Log-Rank / Cox-proportional hazards (pooled and individual OFG doses)	Up to Week 176 visit	Key secondary, main analysis; exploratory analysis
		Efficacy / Randomized participants with prediabetes	Stratified Log-Rank / Cox-proportional hazards model (pooled and individual OFG doses)	Up to Week 176 visit	Key secondary, main analysis; exploratory analysis
		Safety / Randomized participants with prediabetes	Stratified Log-Rank / Cox-proportional hazards model (pooled and individual OFG doses)	Up to Week 190 visit	Key secondary, main analysis; exploratory analysis

	Percentage of participants achieving normoglycemia	Treatment Regimen / Randomized participants with prediabetes	Logistic regression with PMI	Weeks 72,176, and 190 visit	Additional secondary
		Efficacy / Randomized participants with prediabetes	Logistic regression with MI	Weeks 72, 176, and 190 visit	Additional secondary
	Change from baseline in point x OGTT (mg/dL and nmol/L), where x = 30, 60, 90, and 120 minutes	Efficacy	ANCOVA with MI	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM	Weeks 120 and 176 visit	Exploratory analysis
Insulin sensitivity	Change from baseline in Matsuda Index	Efficacy	ANCOVA with MI	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM	Weeks 72 and 176 visit	Exploratory analysis
Fasting insulin	Percent change from baseline in fasting insulin (%)[e]	Efficacy	MMRM (log transformation)	Week 72 visit	Additional secondary
		Treatment Regimen	ANCOVA with PMI (log transformation)	Week 72 visit	Additional secondary
		Efficacy / Randomized participants with prediabetes	MMRM	Week 176 visit	Exploratory analysis
Renal function	Percent change from baseline in UACR (%)	Efficacy	MMRM (log transformation)	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (log transformation)	Weeks 176 visit	Exploratory analysis

	Change from baseline in eGFR CKD-EPI Cystatin-C (ml/min/1.73m^2)	Efficacy	MMRM	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM	Weeks 176 visit	Exploratory analysis
	Change from baseline in eGFR CKD-EPI Creatinine (ml/min/1.73m^2)	Efficacy	MMRM	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM	Weeks 176 visit	Exploratory analysis
	Percent change from baseline in hsCRP (%)	Efficacy	MMRM (log transformation)	Week 72 visit	Exploratory analysis
		Treatment Regimen	ANCOVA with PMI (log transformation)	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (log transformation)	Weeks 176 visit	Exploratory analysis
		Treatment Regimen/ Randomized participants with prediabetes	ANCOVA with PMI (log transformation)		
	Percent change from baseline in FFAs (%)	Efficacy	MMRM (log transformation)	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (log transformation)	Weeks 176 visit	Exploratory analysis
Homeostasis Model Assessment	Percent change from baseline in • HOMA2 B (insulin[e]) • HOMA2 IR (insulin[e])	Efficacy	MMRM (log transformation)	Week 72 visit	Exploratory analysis

	• HOMA2 B (C-peptide) • HOMA2 IR (C-peptide)				
		Efficacy / Randomized participants with prediabetes	MMRM (log transformation)	Weeks 176 visit	Exploratory analysis
	Change from baseline in SF-36v2 Acute Form Physical Functioning domain score	Treatment Regimen / Randomized participants with limitations in physical function at baseline[d]	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Exploratory analysis
		Efficacy / Randomized participants with limitations in physical function at baseline[d]	MMRM (pooled and individual OFG doses)	Week 72 visit	Exploratory analysis
	Change from baseline in SF-36v2 Acute Form domain scores, PCS and MCS	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Additional secondary
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Additional secondary
		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
PGI-S	PGI-S rating	Treatment Regimen	Shift from baseline (pooled and individual OFG doses)	Week 24 and 72 visits	Additional secondary

		Efficacy	Shift from baseline (pooled and individual OFG doses)	Week 24 and 72 visits	Additional secondary
		Treatment Regimen / Randomized participants with prediabetes	Shift from baseline (pooled and individual OFG doses)	Week 120 and 176 visits	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	Shift from baseline (pooled and individual OFG doses)	Week 120 and 176 visits	Exploratory analysis
PGI-C	PGI-C rating	Treatment Regimen	Descriptive statistics (pooled and individual OFG doses)	Week 24 and 72 visits	Additional secondary
		Efficacy	Descriptive statistics (pooled and individual OFG doses)	Week 24 and 72 visits	Additional secondary
		Treatment Regimen / Randomized participants with prediabetes	Descriptive statistics (pooled and individual OFG doses)	Week 120 and 176 visits	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	Descriptive statistics (pooled and individual OFG doses)	Week 120 and 176 visits	Exploratory analysis
IWQOL-Lite-CT	Change from baseline in: • physical function composite score • physical domain score • psychosocial domain score • total score	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Additional secondary
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Additional secondary
		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis

Approved on 25 Jun 2025 GMT

EQ-5D-5L	Change from baseline in EQ-5D-5L utility index and VAS	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 72 visit	Additional secondary
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Additional secondary
		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (pooled and individual OFG doses)	Week 176 visit	Exploratory analysis
PROMIS SD	Individual items	Treatment Regimen	Descriptive statistics	Week 24 and 72 visits	Exploratory analysis
		Efficacy	Descriptive statistics	Week 24 and 72 visit	Exploratory analysis
		Treatment Regimen / Randomized participants with prediabetes	Descriptive statistics	Week 120 and 176 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	Descriptive statistics	Week 120 and 176 visit	Exploratory analysis
	Change from baseline in PROMIS SD T-scores	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 24 and 72 visits	Exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 72 visit	Exploratory analysis
		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI (pooled and individual OFG doses)	Week 120 and 176 visits	Exploratory analysis
		Efficacy / Randomized participants with prediabetes	MMRM (pooled and individual OFG doses)	Week 176	Exploratory analysis
Appetite VAS	Change from baseline in appetite VAS item and overall scores	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 24 and 72 visits	Exploratory analysis

Approved on 25 Jun 2025 GMT

		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI (pooled and individual OFG doses)	Week 176	Exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 24 and 72 visits	Exploratory analysis
		Safety / Randomized participants without prediabetes		Week 74 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes		Week 176 visit	Exploratory analysis
	Change from Week 72 in appetite VAS item and overall scores	Safety / Randomized participants without prediabetes	ANCOVA with MI (pooled and individual OFG doses)	Week 74 visit	Exploratory analysis
Power of food scale (PFS)	Change from baseline in PFS domain and overall scores	Treatment Regimen	ANCOVA with PMI (pooled and individual OFG doses)	Week 24 and 72 visits	Exploratory analysis
		Treatment Regimen / Randomized participants with prediabetes	ANCOVA with PMI (pooled and individual OFG doses)	Week 176	Exploratory analysis
		Efficacy	MMRM (pooled and individual OFG doses)	Week 24 and 72 visits	Exploratory analysis
		Safety / Randomized participants without prediabetes		Week 74 visit	Exploratory analysis
		Efficacy / Randomized participants with prediabetes		Week 176 visit	Exploratory analysis
	Change from Week 72 in appetite VAS item and overall scores	Safety / Randomized participants without prediabetes	ANCOVA with MI (pooled and individual OFG doses)	Week 74 visit	Exploratory analysis

Approved on 25 Jun 2025 GMT

Abbreviations: ANCOVA = analysis of covariance; BMI = body mass index; CDF = cumulative distribution function; CSV = comma separated values; DBP = diastolic blood pressure; eGFR = estimated glomerular filtration rate; FFA = free fatty acids; HDL = high-density lipoprotein; HOMA = homeostasis model assessment; IWQOL-Lite-CT = impact of weight on quality of life-lite clinical trials; hsCRP = high-sensitivity C-reactive protein; LDL = low-density lipoprotein; MCS = Mental Component Score; MI = multiple imputation; MMRM = mixed-effects model for repeated measures; OFG = orforglipron; OGTT = oral glucose tolerance test; PGI-S = patient global impression of severity; PGI-C= patient global impression of change; PCS = Physical Component Score; PFS = power of food scale; PMI = primary multiple imputation; PROMIS SD = patient-reported outcomes measurement information system sleep disturbance;
SBP = systolic blood pressure; SF-36 = Short Form 36-item Health Survey; TP = tipping point; UACR = urine albumin-creatinine ratio; VAS = visual analog scale; VLDL = very low-density lipoprotein.

[a] Population is displayed if different from Randomized Participants.

[b] Assessments collected at multiple postbaseline visits will be analyzed at those scheduled visits using MMRM as supplemental analyses, in addition to the primary time point listed in the table.

[c] PMI strategy will be performed after the log transformation.

[d] The limitation in physical function at baseline is defined as PGI-S response at baseline of “moderately limited,” “very much limited,” or “extremely limited.”

[e] Fasting insulin is collected at the 0 minute OGTT timepoint for the Week 176 analysis.

4.4.1. Key Secondary Endpoints

4.4.1.1. Definition of endpoints

The definitions of key secondary endpoints are specified in Table GZGP.4.5.

4.4.1.2. Main analytical approach

The analytical approaches for the key secondary endpoints are specified in Table GZGP.4.6.

4.4.1.3. Sensitivity Analyses

No sensitivity analyses are planned for key secondary endpoints.

4.4.1.4. Supplementary analyses

No supplemental analyses are planned for key secondary endpoints.

4.4.2. Supportive secondary Endpoints

The definitions of additional secondary endpoints are specified in Table GZGP.4.5. Additionally, the analytical approaches are specified in Table GZGP.4.6.

4.5. Exploratory Endpoints/Estimands Analysis

Exploratory endpoints and analyses are specified in Table GZGP.4.5 and Table GZGP.4.6. Additional exploratory analyses will be described in the Exploratory/Supplemental Analyses Plan and Health Technology Analyses Plan.

4.6. Safety Analyses

The planned safety analyses are consistent with compound-level safety standards, which are based on various sources, including company standards, internal and external subject matter experts, publications from cross-industry initiatives (for example, PHUSE 2013, PHUSE 2015, PHUSE 2017, PHUSE 2018, PHUSE 2022), and publications from regulatory agencies (for example, EMA 2014, CDER/BIRRS 2022a, 2022b). Descriptions of the safety analyses are provided in this SAP, but some details are found in compound-level safety standards.

For safety interpretation, p-values will not be used for hypothesis testing. P-values will be considered as a number between 0 and 1 that gives an idea of how strong the evidence is for an imbalance between study intervention arms. The evidence for an imbalance is stronger toward 0 and weaker toward 1. Similarly, confidence intervals will not be used for hypothesis testing. They reflect the uncertainty of an estimate.

Unless otherwise specified, safety analyses will be conducted using the safety participants (Table GZGP.3.1) and the safety data points set (Table GZGP.3.2). It is noted that for some safety endpoints, such as treatment discontinuation due to AE and so forth, the analyses will be based on data observed during treatment period only.

Summary tables with risk difference will be sorted by decreasing risk difference.

Additionally, the planned safety analyses will be provided for participants with prediabetes at randomization for the final lock.

4.6.1. Extent of Exposure

Duration of exposure to study treatment will be summarized by treatment group for safety participants. Descriptive statistics for participant weeks (or days) of exposure and total participant years in exposure will be provided. Overall exposure will be summarized in total participant-year (PY) of exposure, derived in the following manner:

total PY of exposure = sum of duration of exposure in days (for all participants in treatment group) / 365.25

The frequency of participants falling into different exposure ranges will be summarized:

- >0; ≥4 weeks; ≥8 weeks; ≥12 weeks; ≥16 weeks; ≥20 weeks; ≥24 weeks; ≥36 weeks; ≥48 weeks, ≥60 weeks, and ≥72 weeks.

For participants with prediabetes at baseline, the following additional exposure ranges will be summarized:

- ≥84 weeks; ≥96 weeks; ≥108 weeks; ≥120 weeks; ≥132 weeks; ≥144 weeks; ≥156 weeks; ≥168 weeks, and ≥176 weeks.

No p-values will be reported.

4.6.2. Adverse Events

The planned summaries for AEs are provided in Table GZGP.4.7 and are described more fully in compound-level safety standards.

Table GZGP.4.7. Summary Tables Related to Adverse Events

Analysis	Method	Population/Period
Overview of AEs, including • TEAE • SAE • death, and • permanent discontinuation from study intervention due to an AE.	Fisher's exact	Safety/TP + FP (TP for study treatment discontinuation due to an AE)
TEAEs by PT within SOC	Fisher's exact	Safety/TP + FP
TEAEs by PT	Fisher's exact	Safety/TP + FP
Maximum Severity TEAEs by PT within SOC		Safety/TP + FP
TEAEs with incidence ≥2% by PT	Fisher's exact	Safety/TP + FP
SAEs by PT within SOC	Fisher's exact	Safety/TP + FP
Primary AEs leading to permanent discontinuation of study intervention by PT within SOC	Fisher's exact	Safety/TP
Primary AEs leading to permanent discontinuation of study by PT within SOC	Fisher's exact	Safety/TP + FP
AEs leading to study intervention interruption by PT within SOC	Fisher's exact	Safety/TP
AEs leading to study intervention reduction by PT within SOC	Fisher's exact	Safety/TP
Listing of SAEs		Safety/TP + FP

Analysis	Method	Population/Period
Listing of primary AEs leading to permanent discontinuation of study intervention		Safety/TP
Listing of primary AEs leading to permanent discontinuation of study		Safety/TP + FP
Listing of deaths		Safety/TP + FP
Listing AEs related to suspected overdosing		Safety/TP
Listing of participants with at least 1 notable event		Randomized/TP + FP

Abbreviations: AE =Adverse Event; FP = follow-up period; TP = treatment period; TEAE = Treatment Emergent Adverse Event; SAE = Serious Adverse Event; PT = Preferred Term; SOC = System Organ Class.

4.6.3. Clinical Laboratory Evaluation

The planned summaries for clinical laboratory evaluations are provided in Table GZGP.4.8 and are described more fully in compound-level safety standards.

Table GZGP.4.8. Summary Tables Related to Clinical Laboratory Evaluations

Analysis	Method	Population/Period
Box plots and mean/SD (or 95% CI) for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plots and mean/SD (or 95% CI) for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Summary for participants with elevated or low values meeting specified levels	Descriptive statistics	Safety/TP + FP
Listing of abnormal laboratory findings		Safety/TP + FP

Abbreviations: CI = confidence intervals; FP = follow-up period; SD = standard deviation; TP = treatment period.

4.6.4. Vital Signs and Physical Characteristics

Triplicate vital signs will be collected at each visit, thus the mean of these measurements will be used for the vital signs analyses. The planned summaries for vital signs (SBP, diastolic blood pressure, and pulse rate) are provided in Table GZGP.4.9, and are described more fully in the compound-level safety standards.

Table GZGP.4.9. Summary Tables Related to Vital Signs

Analysis	Method	Population/Period
Box plots and mean/SD (or 95% CI) for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plots and mean/SD (or 95% CI) for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Analysis of blood pressure and pulse rate for change from baseline	MMRM	Safety/TP + FP
Summary for participants meeting specific blood pressure and pulse rate levels	Descriptive statistics	Safety/TP + FP
Shift of maximum-to-maximum for pulse rate	Descriptive statistics	Safety/TP + FP

Abbreviations: CI = confidence intervals; FP = follow-up period; MMRM = mixed model repeated measures; SD = standard deviation; TP = treatment period.

4.6.5. Electrocardiograms

Triplicate electrocardiograms (ECGs) will be collected at the same visit; thus the mean of these measurements will be used for the analysis. The planned summaries for ECG parameters are

provided in Table GZGP.4.10 and are described more fully in the compound-level safety standards.

Table GZGP.4.10. Summary Tables Related to ECG Parameters

Analysis	Method	Population/Period
Box plot and mean/SD (or 95% CI) for observed values by visit	Descriptive statistics	Safety/TP + FP
Box plot and mean/SD (or 95% CI) for change from baseline values by visit	Descriptive statistics	Safety/TP + FP
Heart rate and PR interval in specified categories	Descriptive statistics	Safety/TP + FP
Analysis for change from baseline in ECG parameters (heart rate, PR interval)	MMRM	Safety/TP + FP

Abbreviations: CI = confidence intervals; FP = follow-up period; MMRM = mixed model repeated measures; SD = standard deviation; TP = treatment period.

4.6.6. Safety Topics of Interest

This section includes safety topics of interest whether due to observed safety findings, potential findings based on drug class, or safety topics anticipated to be requested by a regulatory agency for any reason. In general, safety topics of interest will be identified by one or more standardized Medical Dictionary for Regulatory Activities (MedDRA) query(ies) (SMQs), system organ class (SOC), high level terms (HLT), FDA medical query(ies) (FMQ), or by a Lilly-defined MedDRA preferred term (PT) listing based upon the review of the most current MedDRA Version, or by relevant laboratory changes. Search criteria are detailed in the compound-level safety standards.

The planned analyses for safety topics of interest are provided in Table GZGP.4.11 and are described more fully in compound-level safety standards.

Table GZGP.4.11. Description and Analyses of Safety of Interest

Special Safety Topic	Short Description	Analysis	Method	Population/Period
Major Adverse Cardiovascular Events	Death and nonfatal CV AEs will be adjudicated by a committee of physicians external to Lilly with cardiology expertise: Clinical Endpoint Committee (CEC).	Positively adjudicated MACE by category/subcategory MACE reported by investigators may be summarized		Safety/TP + FP
		Listing of all MACE reported by investigator (whether or not positively adjudicated)		Safety/TP + FP
Arrhythmias and Cardiac Conduction Disorders	The treatment-emergent (TE) arrhythmias and cardiac conduction disorders events will be derived using the MedDRA PTs contained in certain SMQs and HLT.	TE arrhythmias and cardiac conduction disorders by PT nested within SMQ and HLT[a]	Fisher's exact	Safety/TP + FP
Hypotension, Orthostatic Hypotension, and Syncope	The TE hypotension, orthostatic hypotension, and syncope events will be derived using MedDRA PTs and FMQs.	TE hypotension, orthostatic hypotension, and syncope by PT[a]	Fisher's exact	Safety/TP + FP
		Hypotension by baseline antihypertensive medication use using hypotension narrow FMQ		
Hypoglycemia[a,b]	The American Diabetes Association position statement on glycemic targets (ADA 2025) will be used to define: • Level 1 hypoglycemia • Level 2 hypoglycemia • Level 3 hypoglycemia (severe/serious hypoglycemia) Nocturnal hypoglycemia events (including severe hypoglycemia) **occur at night** and presumably during sleep.	Incidence (percent of patients experiencing 1 or more episode) of level 2 or level 3 hypoglycemia	Logistic regression for the overall postbaseline period	Safety/TP + FP excluding events after initiation of new antihyperglycemic therapy as defined in Section 6.3. Supportive analysis: Safety/TP + FP regardless of initiation of new antihyperglycemic therapy
		Incidence of level 3 (severe/serious) hypoglycemia		
		Incidence of level 1 hypoglycemia (if warranted by data) [a]		
		Rate (episodes/patient/year) of level 2 or level 3 hypoglycemia	Empirical method for the overall postbaseline period	
		Rate of level 3 (severe/serious) hypoglycemia		
		Rate of level 1 hypoglycemia (if warranted by data)		
		Listing of level 2 or level 3 hypoglycemia events		
		Listing of level 2 or level 3 nocturnal hypoglycemia events		Safety/TP + FP
		Severe or serious TE gastrointestinal events by PT	Fisher's exact	Safety/TP + FP

Gastrointestinal (GI) Safety[c]	GI AEs using Gastrointestinal disorders SOC and FMQ will be captured.	Study intervention discontinuation due to TE gastrointestinal events	Fisher's exact	Safety/TP + FP
		TE abdominal pain by FMQ narrow PT	Fisher's exact	Safety/TP + FP
		Prevalence and incidence over time for TE nausea, vomiting, diarrhoea, constipation, and NVD		Safety/TP + FP
		Plot of time to the onset of TE nausea, vomiting, diarrhoea, constipation, and NVD	KM	Safety/TP + FP
		Plot of prevalence and incidence over time for TE nausea, vomiting, diarrhoea, constipation, and NVD by maximum severity		Safety/TP + FP
Renal Safety	Laboratory measures related to renal safety will be analyzed. Renal events including acute renal failure and chronic renal failure exacerbation will be captured using SMQs. Dehydration events will be captured using SMQ.	Shift of minimum-to-minimum for eGFR as estimated by the CKD-EPI Cystatin C equation		Safety/TP + FP
		Shift of maximum-to-maximum for UACR		Safety/TP + FP
		Summary of eGFR decrease from baseline		Safety/TP + FP
		MMRM analyses for eGFR estimated by the CKD-EPI Cystatin C equation	MMRM	Safety/TP + FP
		MMRM analyses for UACR (log transformation)	MMRM	Safety/TP + FP
		TE renal events by PT nested within SMQ [a]	Fisher's exact	Safety/TP + FP
		Listing of TE dehydration events by PT [a]		Safety/TP + FP
Pancreatic Safety	The pancreatic enzyme data (p-amylase and lipase) will be observed through laboratory testing. All suspected cases of acute or chronic pancreatitis and AEs of severe or serious abdominal pain of unknown etiology will be sent for adjudication by an independent clinical endpoint committee.	Shift of maximum-to-maximum for pancreatic enzyme		Safety/TP + FP
		MMRM analysis for pancreatic enzymes (p-amylase and lipase) with a log transformation (postbaseline/baseline)	MMRM	Safety/TP + FP
		Investigator-reported events and subsequently confirmed adjudicated events, respectively	Fisher's exact	Safety/TP + FP
		TE Pancreatic events by "Acute pancreatitis" SMQ (narrow scope) and "Chronic pancreatitis" PT		Safety/TP + FP
		Listing of adjudicated and investigator-reported pancreatic events		Safety/TP + FP
Thyroid Safety		TE thyroid C-cell hyperplasia and malignancies by PT within Event Category	Fisher's exact	Safety/TP + FP

	TE thyroid malignancies and C-cell hyperplasia will be identified using MedDRA HLT and PT. The purpose of calcitonin measurements is to assess the potential effect of orforglipron on thyroid C-cell function.	Shift of maximum-to-maximum for calcitonin value in the thresholds		Safety/TP + FP
		Listing of participants who meet protocol defined discontinuation criteria based on calcitonin: • with eGFR <60 mL/min/1.73 m2 at baseline, serum calcitonin value ≥35 ng/L AND ≥50% increase from the baseline value[a] • with eGFR ≥60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥20 and <35 ng/L AND ≥50% increase from the baseline value and ≥10% increase on retest[a] • with eGFR ≥60 mL/min/1.73 m^2 at baseline, serum calcitonin value ≥35 ng/L AND ≥50% increase from the baseline value[a]		Safety/TP + FP
		MMRM analysis for calcitonin (log transformation)	MMRM	Safety/TP + FP
Malignancies	The malignancy events will be derived using the MedDRA PTs contained certain SMQs.	TE malignancy by PT nested within SMQ[a]	Fisher's exact	Safety/TP + FP
Hepatic Safety	Hepatic labs include ALT, AST, ALP, TBL, DBL, and GGT. When criteria are met for hepatic evaluations, investigators will conduct close monitoring of hepatic symptoms and liver tests, perform a comprehensive evaluation for alternative causes of abnormal liver tests, and complete follow-up hepatic safety eCRFs. Potentially drug-related hepatic disorders will be identified using SMQs.	Abnormal postbaseline categories for hepatic safety parameters: ALT, AST, ALP, TBL, DBL, and GGT		Safety/TP + FP
		Treatment-emergent potentially drug-related hepatic disorders by PT nested within SMQ	Fisher's exact	Safety/TP + FP
		Hepatocellular drug-induced liver injury screening plot (maximum TBL vs maximum ALT or AST)		Safety/TP + FP
		Hepatocellular drug-induced liver injury screening table		Safety/TP + FP
		Cholestatic drug-induced liver injury screening table		Safety/TP + FP
		Cholestatic drug-induced liver injury screening plot (maximum TBL vs maximum ALP)		Safety/TP + FP
		Hepatic Laboratory Parameters (ALT, AST, ALP, TBL, DBL, and GGT) with a log transformation (postbaseline/baseline)	MMRM	Safety/TP + FP

		Listing of participants with ALT or AST ≥3X ULN		Safety/TP + FP
		Listing of participants with ALP or TBL ≥2X ULN		Safety/TP + FP
		Shift of maximum-to-maximum for ALT, AST, ALP, TBL		Safety/TP + FP
		Participant profiles for participants meeting criteria for a comprehensive hepatic evaluation (as defined in the protocol).		Safety/TP + FP
Gallbladder and Biliary Tract Disorders	All events of TE gallbladder and biliary tract disorders will be identified by using certain SMQs.	TE gallbladder and biliary tract disorders by PT nested within SMQ [a]	Fisher's exact	Safety/TP + FP
Hypersensitivity Reactions	All events of TE allergic reaction and hypersensitivities will be identified by using certain SMQs.	TE allergic reaction and hypersensitivities by PT nested within SMQ [a]	Fisher's exact	Safety/TP + FP
Depression, Suicidal Ideation, and Behavior	AEs will be searched using MedDRA PTs that satisfy the search criteria.	TE major depressive disorder, suicidal ideation, or behavior events by PT nested within applicable FMQ or SMQ [a]	Fisher's exact	Safety/TP + FP
	Suicide-related thoughts and behaviors will be collected based on the C-SSRS.	Summary of C-SSRS categories and composite measures		Safety/TP + FP
		Listing of C-SSRS categories and composite measures		Safety/TP + FP
	Patient health questionnaire-9 (PHQ-9) will be collected to assess the specific diagnostic symptoms that determine the presence of a clinical depressive disorder. The PHQ-9 total scores will be categorized as minimal to none, mild, moderate, moderately severe, and severe	Shift of each baseline category (maximum value) versus each postbaseline category (maximum value)		Safety/TP + FP
		Summary of categories based on the maximum values during baseline and postbaseline: • any increase in depression category (that is, worsening of depression) • increase from Minimal to none or Mild depression to Moderate, Moderately severe, or Severe depression • increase from Mild or Moderate depression to Moderately severe or Severe depression		Safety/TP + FP
Abuse Potential	AEs will be searched using a modified study agent abuse potential FMQ.	TE abuse potential events by PT [a]	Fisher's exact	Safety/TP + FP

AEs possibly related to loss of lean muscle mass	AEs will be identified using MedDRA HLTs and PTs	TEAEs possibly related to loss of lean muscle mass by PT within Event Category	Fisher's exact	Safety/TP + FP
Malnutrition	AEs will be identified using MedDRA HLTs and PTs	TE malnutrition by PT within Event Category	Fisher's exact	Safety/TP + FP
Excessive weight loss	AEs will be identified using MedDRA PTs BMI <18.5 kg/m2 will be analyzed	Severe or serious TE excessive weight loss by PT or BMI <18.5kg/m^2	Fisher's exact	Safety/TP + FP
		Listing of participants with BMI <18.5 kg/m^2		

Abbreviations: AE = adverse event; ALP = serum alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; BMI = body mass index; CEC = clinical endpoint committee; CKD-EPI = Chronic Kidney Disease Epidemiology; C-SSRS = Columbia-Suicide Severity rating scale; CV = cardiovascular; eCRF = electronic case report form; eGFR = estimated glomerular filtration rate; DBL = direct bilirubin; FMQ = FDA Medical Query; FP = follow-up period; GGT = gamma glutamyl transferase; GI = gastrointestinal; HLT = high-level term; MACE = major adverse cardiovascular event; MedDRA = Medical Dictionary for Regulatory Activities; MMRM = mixed model repeated measure; PT = Preferred Term; SMQ = standardized MedDRA query; SOC = system organ class; TEAE = treatment-emergent adverse event; TBL = total bilirubin; TP = treatment period; UACR = urinary albumin-to-creatinine ratio; vs = versus.

[a] For these tables, if the number of events is low, a listing might be provided instead.

[b] Glucose values collected on the same date as an OGTT assessment are not included

[c] Additionally, all GI events will be analyzed.

Note: Listings and participant profiles may be provided through interactive display tools instead of a static display.

4.7. Other Analyses

4.7.1. Subgroup analyses

Subgroup analyses will be conducted for the primary endpoint with the treatment regimen estimand:

- percent change in body weight from baseline at Week 72 visit.

The variables for subgroup analysis are specified in Table GZGP.6.1.

The ANCOVA model specified in Section 4.1.2.1.1, will be fitted separately within each subgroup. The least squares (LS) mean, LS mean difference, standard error (SE), and 95% confidence interval will be presented. The LS means and variance-covariance estimates from these separate models will be used to test the treatment-by-subgroup interaction at a significance level of 0.10.

If the subgroup factor is part of the strata variable, a new strata variable that excludes the subgroup categories will be used.

If any category within the subgroup is <5% of the total population, only descriptive statistics will be provided for that category (that is, there will be no inferential testing within the subgroup category).

A forest plot including the treatment difference and 95% confidence interval estimated for each subgroup level will be presented.

A Bayesian shrinkage method will be used to provide adjusted estimates and improve estimation precision. Following the paper by Wang et. al. (2024), for a given baseline characteristic with k subgroups, let Y_i $(i = 1, \ldots, k)$ be the observed sample estimate of the treatment effect difference in subgroup i. The following hierarchical model will be used:

$$Y_i \sim N(\mu_i, \sigma_i^2),$$

$$\mu_i \sim N(\mu, \tau^2),$$

$$\mu \sim N(0, 30^2),$$

$$\tau \sim N^+(10),$$

where σ_i^2 is set to the observed variance for sample estimate, $N^+(10)$ is the half-normal distribution with a scaled parameter of 10. A weakly informative prior, as suggested by Röver et al. (2021) and specified above, will be applied to μ and τ. Of note, the unit information standard deviation (UISD) is assumed to be 15% based on historical weight management studies, and a standard deviation of 30% is chosen for the centrality parameter μ , so that it is approximately twice the unit information standard deviation (UISD). A scale parameter of 10 is chosen for the half-normal distribution such that the median heterogeneity in percent body weight reduction is approximately 6.7%, which is a reasonably large difference.

Additional subgroup evaluations may be conducted as exploratory analyses.

4.7.2. Supplementary Analyses

A sensitivity supplementary analysis will be conducted for all primary and key secondary efficacy endpoints based on the population specified in Section 3, excluding participants randomized at sites closed due to Good Clinical Practice (GCP) non-compliance issues. Each endpoint will be evaluated based on the approaches described in Sections 4.3.2 and 4.4.1.2. For safety analyses, overview of adverse events and summary of treatment-emergent adverse events will be conducted, also excluding participants randomized at sites closed due to GCP non-compliance issues.

4.8. Interim Analyses

No interim analyses are planned for Study GZGP. If an unplanned interim analysis is deemed necessary for reasons other than a safety concern, Protocol GZGP must be amended.

4.8.1. Data Monitoring Committee

A data monitoring committee (DMC) is not planned for this study.

An independent DMC will be established for interim safety monitoring of Study J2A-MC-GZGS (ACHIEVE-4) (GZGS), "A Phase 3, Open Label, Study of Once-Daily orforglipron Compared with Insulin Glargine in Adult Participants with Type 2 Diabetes and Obesity or Overweight at Increased Cardiovascular Risk", a study within the orforglipron Phase 3 program for chronic weight management (CWM) and T2D. The DMC will have the responsibility to review unblinded interim safety analysis results from Study GZGS. The DMC may be asked to review unblinded safety data from Study GZGP if a need arises following the blinded trial-level safety reviews (TLSRs) conducted by the sponsor. If needed, permanent data transfers will occur for the purposes of supporting the DMC. Only the statisticians from the statistical analysis center (SAC) will have access to the unblinded data that are presented to the DMC. The SAC and members of the DMC will abide by the principles and responsibilities described in the Study GZGS DMC charter, which includes keeping all unblinded information confidential until the planned unblinding of the trial.

4.9. Changes to Protocol-Planned Analyses

There are changes to the analyses described in the protocol:

1. orforglipron 6 mg endpoint "percentage of participants who achieve a body weight reduction of ≥20%" was removed from the key secondary objectives, and added to additional secondary objectives.
2. treatment compliance is defined as taking at least 75% of required study intervention during the treatment period.

5. Sample Size Determination

A sample size of 3,042 participants (702 participants per orforglipron treatment group and 936 participants in the placebo group) provides more than 90% power to demonstrate superiority of 6 mg, 12 mg, and/or 36 mg orforglipron to placebo with regards to mean percent change in body weight from baseline to Week 72. The sample size determination assumes that evaluation of superiority of 6 mg, 12 mg, and 36 mg orforglipron to placebo will be conducted in parallel, each at a 2-sided significance level of 0.016 using a 2-sample t-test for the treatment regimen estimand. Additionally, a difference of at least 5% mean body weight reduction from baseline at 72 weeks for 6 mg, 12 mg, and 36 mg orforglipron compared with placebo, a common standard deviation (SD) of 10%, and a dropout rate of 30% for placebo and 20% for the orforglipron treatment groups are assumed for the statistical power calculation.

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics

Demographics and baseline characteristics will be summarized by treatment group for randomly assigned participants. The listing of basic demographic characteristics (that is, age, sex, ethnicity, race, country, baseline body weight, and so forth.) for randomly assigned participants will be provided. Additionally, demographics and baseline characteristics will be provided for participants with prediabetes at randomization for the final lock.

The continuous variables will be summarized using descriptive statistics and the categorical variables will be summarized using frequency counts and percentages.

Table GZGP.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 4.7.1.

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

Variable	Quantitative Summary	Categorical Summary	Subgroup Analysis[a]
Demographics			
Age[b]	Yes	<65, ≥65 years	X
		<75, ≥75 years	X
		<65, ≥65 and <75, ≥75 and <85, ≥85 years	
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	X
Geographic region	No	North America, Europe, Asia, or Central/South America	X
Height (cm)	Yes		
Baseline waist circumference (cm)	Yes		
Baseline body weight (kg)	Yes		
Baseline BMI	Yes	≥27 to <30, ≥30 to <35, ≥35 to <40, ≥40 kg/m^2	X
		≤ median, > median	
Caffeine use	No	Never, Current, Former	
Alcohol use	No	Never, Current, Former	
Tobacco use	No	Never, Current, Former	

Nicotine Replacement use	No	Never, Current, Former	
Recreational drug use - methamphetamine	No	Never, Current, Former	
Recreational drug use - cocaine	No	Never, Current, Former	
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 Cystatin-C (mL/min/1.73m^2)	Yes	<60, ≥60 mL/min/1.73m^2	
Baseline eGFR CKD-EPI Creatinine (mL/min/1.73m^2)	Yes	<60, ≥60 mL/min/1.73m^2	
Baseline UACR (g/kg)	Yes	<30, ≥30 and ≤300, >300 g/kg	
Baseline triglycerides (mg/dL)	Yes		
Baseline total cholesterol (mg/dL)	Yes		
Baseline VLDL-cholesterol (mg/dL)	Yes		
Baseline non-HDL-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		
Baseline HbA1c (mmol/mol)	Yes		
Baseline fasting glucose (mg/dL)	Yes		
Baseline fasting glucose (mmol/L)	Yes		
Duration of obesity (years)	Yes		
Prediabetes[c]	No	Yes, No	X

Abbreviations: eGFR = estimated glomerular filtration rate; HDL = high-density lipoprotein; LDL = low-density lipoprotein; UACR = urine albumin-to-creatinine ratio; VLDL = very low-density lipoprotein.

[a] Subgroup analyses are defined in Section 4.7.1 with more details.

[b] Age in years will be calculated as length of the time interval from the imputed date of birth (July 1 in the year of birth collected in the eCRF) to the informed consent date.

[c] Prediabetes status at baseline is determined from the laboratory data in LabsConnect.

6.2. Appendix 2: Historical Illnesses and Preexisting Conditions

Historical illness is defined as a condition/event with an end date prior to the date of informed consent.

A preexisting condition is defined as a condition/event with a start date prior to the first dose of the study intervention and stop date that is at or after the informed consent date or has no stop date (that is, ongoing). Randomization date will be used if a participant has not been dosed.

The planned summaries for historical illness and preexisting conditions are provided in Table GZGP.6.2. Additionally, analyses for historical illnesses and preexisting conditions will be provided for participants with prediabetes at randomization for the final lock.

Table GZGP.6.2. Summary Tables Related to Historical Illness and Preexisting Conditions

Analysis	Population/Period
Historical illness by PT within SOC	Randomized
Preexisting conditions by PT within SOC	Randomized
Obesity related health disorders[a]	Randomized
Comorbidities at baseline[b]	Randomized

Abbreviations: PT = preferred term; SOC = system organ class.

[a] Include the 11 Obesity-Related Health Disorders defined in Table GZGP.6.6.

[b] Include hypertension, dyslipidemia, coronary artery disease, cerebrovascular disease, obstructive sleep apnea, osteoarthritis, anxiety/depression, metabolic dysfunction-associated steatotic liver disease, asthma or chronic obstructive pulmonary disease, gout or hyperuricemia, renal disease, and female reproductive disorders.

6.3. Appendix 3: Prior/Concomitant Medications

Medications that start before or at the last date of treatment period or follow-up period and are ongoing or ended during the treatment period or follow-up period will be classified as concomitant medication. Medications that start and end before the first dose date of the study intervention will be classified as prior therapy.

Baseline is defined as the corresponding medication taken on the day before the date of the first dose of study intervention.

If there are no doses of study intervention, randomization date will be used instead of the first dose date.

The planned summaries for concomitant medications are provided in Table GZGP.6.3. Additionally, analyses for concomitant medications will be provided for participants with prediabetes at randomization for the final lock.

Table GZGP.6.3. Summary Tables Related to Concomitant Medications

Analysis		Population/Period
CMs by PN		Randomized/TP + FP
CMs by PN within ATC code		Randomized/TP + FP
Antihypertensive CMs at baseline by PN within ATC code		Randomized
Lipid lowering CMs at baseline by PN within ATC code		Randomized
Status change of antihypertensive and lipid lowering CMs		Randomized/TP
Antihyperglycemic CM initiated after baseline by PN within ATC code		Randomized/TP
Antidiarrheal, Antiemetic, and Constipation CMs at baseline by PN within ATC code		
Antidiarrheal, Antiemetic, and Constipation CMs initiated after baseline by PN within ATC code		Randomized/TP
Prohibited concomitant medication and surgical procedure on weight management treatment by PN within ATC code		Randomized Participants/TP
Prohibited concomitant medication on anti-hyperglycemic therapy by PN within ATC code		Randomized Participants/TP
Obesity management medications 1 year prior to screening by PN within ATC code		Randomized

Abbreviations: ATC = Anatomical Therapeutic Chemical; CM = concomitant medication; PN = preferred name; FP = follow-up period; TP = treatment period.

Table GZGP.6.4 provides the protocol-specified concomitant medications of interest.

Table GZGP.6.4. Protocol-Specified Concomitant Medication

Category	Preferred Name / ATC Code
Antihypertensive Medication	All medications with ATC code containing ‘C01’, ‘C03’, ‘C07’, ‘C08’, or ‘C09’
Lipid Lowering Medication	All medications with ATC code containing ‘C10’
Antihyperglycemic Medication[a]	All medications with ATC code containing ‘A10', except for ‘A10XA'

Abbreviations: ATC = Anatomical Therapeutic Chemical; T2D = type 2 diabetes.

[a] For participants who develop T2D during the conduct of the study.

Table GZGP.6.5 provides the protocol-specified prohibited concomitant medication and surgical procedure rules.

Table GZGP.6.5. Prohibited Concomitant Medication Rules

Timing	Preferred Name / ATC Code	ATC Classification	Rule
Prohibited Concomitant Medication on Weight Management Treatment[a]			
Post baseline	All medications with ATC code of A08AA	Centrally acting antiobesity products	Any initiation is prohibited during treatment period
	All medications with ATC code of A08AB	Peripherally acting antiobesity products	Any initiation is prohibited during treatment period
	All medications with ATC code of A08AX	Other antiobesity drugs	Any initiation is prohibited during treatment period
	All medications with ATC code A08AW	Herbal antiobesity preparations	Any initiation is prohibited during treatment period
	All medications that contain phenylpropanolamine, tirzepatide, semaglutide, or liraglutide in the preferred term		Any initiation is prohibited during treatment period
Prohibited Surgical Procedure on Weight Management Treatment[a]			
Post-baseline	ABDOMINOPLASTY		Any initiation is prohibited during treatment period
	GASTRIC BANDING		Any initiation is prohibited during treatment period
	GASTRIC BYPASS		Any initiation is prohibited during treatment period
	LIPOSUCTION		Any initiation is prohibited during treatment period
	SLEEVE GASTRECTOMY		Any initiation is prohibited during treatment period
	CRYOLIPOLYSIS		Any initiation is prohibited during treatment period
	MUCOSAL ABLATION		Any initiation is prohibited during treatment period
	GASTRIC ARTERY EMBOLIZATION		Any initiation is prohibited during treatment period
	INTRAGASTRIC BALLOON		Any initiation is prohibited during treatment period
	DUODENAL-JEJUNAL ENDOLUMINAL LINER		Any initiation is prohibited during treatment period
Prohibited Concomitant Medication on Anti-hyperglycemic Therapy			
Post-baseline	All medications with ATC code of A10BH	Dipeptidyl peptidase 4 (DPP-4) inhibitors	Any initiation is prohibited during treatment period

Timing	Preferred Name / ATC Code	ATC Classification	Rule
	All medications with ATC code of A10BJ	Glucagon-like peptide-1 (GLP-1) analogues	Any initiation is prohibited during treatment period
	All medications with ATC code of A10BX	Other blood glucose lowering drugs, excluding insulins	Any initiation is prohibited during treatment period
	All medications that contain tirzepatide, semaglutide, or liraglutide in the preferred term		Any initiation is prohibited during treatment period
	All medications with ATC code A10BD and preferred term containing any of the following: 'liptin' 'tide'	Combinations of oral blood glucose lowering drugs	Any initiation is prohibited during treatment period
	All medications containing the following in the preferred term: Empagliflozin, Dapagliflozin, Canagliflozin, Tofogliflozin	Sodium-glucose co-transporter 2 (SGLT2) inhibitors	Any initiation prior to the diagnosis of T2D during treatment period
	All medications with ATC code A10BA	Biguanides	Any initiation prior to the diagnosis of T2D during treatment period
	All medications with ATC code A10BD and preferred term containing any of the following: 'Metformin' 'gliflozin'	Combinations of oral blood glucose lowering drugs	Any initiation prior to the diagnosis of T2D during treatment period

Abbreviations: ATC = Anatomical Therapeutic Chemical; T2D = type 2 diabetes; ICE = intercurrent event.

[a] Use of these medications will be reviewed by the study team (blinded to study treatment) on a case-by-case basis to assess the clinical significance of the change in medication. After review, if deemed to be non-clinically significant, the case won't be considered as an ICE.

6.4. Appendix 4: Treatment Compliance

Treatment compliance is defined as taking at least 75% of required study intervention during the treatment period. Compliance will be calculated by

([total number of doses dispensed – total number of doses returned]/total number of doses expected to be administered) × 100%.

Frequency counts and percentages of participants compliant to study intervention will be summarized by treatment group using the safety participants during treatment period. No inferential analysis will be performed.

Additionally, analyses for treatment compliance will be provided for participants with prediabetes at randomization for the final lock.

Participants with dose interruptions/modifications will be summarized with reasons. Interruptions will be summarized if the duration of the interruption is 7 days or longer for all reasons other than an AE.

6.5. Appendix 5: Important Protocol Deviations

Important protocol deviations are identified in the Trial Issues Management Plan. A listing and summary of important protocol deviations by treatment group will be provided for all randomly assigned participants. No inferential analysis will be performed. Additionally, analyses for important protocol deviations will be provided for participants with prediabetes at randomization for the final lock.

6.6. Appendix 6: Clinical Trial Registry Analyses

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

Analyses provided for the CTR requirements include the following

Summary of AEs, provided as a dataset which will be converted to an XML file. Both serious AEs (SAEs) and "Other" Non-Serious Adverse Events are summarized by treatment group and MedDRA PT.

- An AE is considered "Serious" whether or not it is a TEAE.
- An AE is considered in the "Other" category if it is both a TEAE and is not serious. For each SAE and "Other" AE, for each term and treatment group, the following are provided
    - the number of participants at risk of an event
    - the number of participants who experienced each event term, and
    - the number of events experienced.
- For each SAE, these additional terms are provided for EudraCT:
    - the total number of occurrences causally related to treatment
    - the total number of deaths, and
    - the total number of deaths causally related to treatment.
- Consistent with www.ClinicalTrials.gov requirements, "Other" AEs that occur in fewer than 5% of participants in every treatment group may be excluded if a 5% threshold is chosen. Allowable thresholds include 0% (all events), 1%, 2%, 3%, 4% and 5%.
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

Demographic table including the following age ranges required by EudraCT: ≥18 to <65 years, ≥65 to <85 years, and ≥85 years.

6.7. Appendix 7: Body Composition Assessments via Dual-Energy X-ray Absorptiometry (DXA)

This section is applicable to the participants who are enrolled in the Dual Energy X-ray Absorptiometry (DXA) appendix.

The primary objective of DXA addendum is to demonstrate that orforglipron (6 mg, 12 mg, and 36 mg pooled doses) once daily (QD) is superior to placebo in percent change in total body fat mass loss from baseline to 72 weeks.

Percent change of total body fat mass will be calculated as

[(Total body fat mass at 72 weeks - Total body fat mass at baseline) / Total body fat mass at baseline] ×100

The secondary objectives include:

- To demonstrate that orforglipron (6 mg, 12 mg, and 36 mg pooled doses) QD is superior to placebo with regards to change in total body fat mass (kg) from baseline to 72 weeks, and
- To assess, for orforglipron (6 mg, 12 mg, and 36 mg pooled doses) QD, the following parameters, from baseline to 72 weeks:
    - percent change in total body lean mass, and
    - change in total body lean mass (kg).

Percent change of total body lean mass will be calculated as

[(Total body lean mass at 72 weeks - Total body lean mass at baseline) / Total body lean mass at baseline] × 100

DXA will be performed at the baseline and at the end of 72-week treatment (Visit 21), or at the early discontinuation visit (ED).

In addition to the endpoints listed above, the following parameters may also be analyzed:

- total body mass (kg)
- total body fat mass (kg)
- total body lean mass (kg)
- percent body fat mass = total body fat mass / total body mass × 100
- percent body lean mass = total body lean mass / total body mass × 100
- fat lean mass ratio = total body fat mass / total body lean mass
- visceral fat mass in android region (kg)
- percent visceral fat mass = visceral fat mass in android region / (body fat mass in android region + body lean mass in android region) × 100
- fat loss index (FLI) = total body fat mass change / (total body fat mass change + total body lean mass change) × 100

Unless otherwise specified, all analyses for the variables measured or derived from DXA will be conducted on all randomized participants who have a baseline DXA scan. Baseline is defined as the last non-missing data collected at randomization (prior to first dosing of study drug).

Descriptive summary statistics (for example, sample size, mean, SD, minimum, maximum, and median) of all the parameters listed above (but not limited to) will be provided by treatment group (placebo, orforglipron 6 mg, orforglipron 12 mg, orforglipron 36 mg, and all orforglipron pooled doses) at baseline and postbaseline visits (for both the actual value and change from baseline value).

In addition, a summary of demographics and baseline characteristics for participants in the DXA addendum will be provided.

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and the confidence interval will be calculated at 95%, 2-sided. There will be no multiplicity adjustment.

The analysis of covariance (ANCOVA) model described in Section 4.1.2.1.1 will be used to analyze the continuous outcomes with treatment regimen estimand data points set. Missing data will be imputed as described in Section 4.1.2.1.2. DXA total body endpoints will be analyzed including and excluding head.

For the FLI metric, least square means (LSMs) from ANCOVA models for total body fat mass change and total body lean mass change will be used to estimate mean FLI in each treatment group. We propose to use LSMs instead of fitting regression models on subject-level data because FLI is subject to wide fluctuations when the denominator is small, and subject-level FLI can have large variability if the body weight reduction is small. Treatment comparison in mean FLI may be performed by using delta method. Median and interquartile range of subject-level FLI in each treatment group may be reported.

A sample size of approximately 156 randomly assigned participants (approximately 39 participants per arm) was planned for this appendix to assess the difference between orforglipron 6 mg, 12 mg, and 36 mg pooled and placebo in the percent change of total fat mass from baseline at 72 weeks. Assuming the SD of the percent change from baseline in total fat mass is 12.4% and a dropout of 30%, a total of 108 participants completing 72 weeks (27 participants in each orforglipron dose group and the placebo group) will provide at least 90% power to detect a statistically significant difference of 9% using a 2-sided t-test and an alpha level of 0.05.

6.8. Appendix 8: Sample Size Determinations for Patient-Reported Outcomes of Cravings and Psychological Impact of Food

A sample size of 94 participants (22 participants per orforglipron treatment group and 28 participants in the placebo group) will provide more than 70% power to demonstrate superiority of 6 mg, 12mg, 36 mg, and/or pooled doses of orforglipron to placebo with regards to mean change from baseline in overall appetite visual analog scale (VAS) score to Week 72. The sample size determination assumes that evaluation of superiority 6 mg, 12 mg, and 36 mg orforglipron to placebo will be conducted each at a 2-sided significance level of 0.05 using a

2-sample t-test. Additionally, a difference of at least 15 mm in mean change from baseline in overall appetite VAS score at Week 72 for 6 mg, 12 mg, 36 mg, and pooled doses of orforglipron compared with placebo, and a common SD of 20 mm (Sadoul 2014).

Additionally, the sample size provides greater than 90% power to demonstrate superiority of 6 mg, 12mg, 36 mg, and/or pooled doses of orforglipron to placebo with regards to mean change from baseline in Power of Food Scale (PFS) score to Week 72. The sample size determination assumes that evaluation of superiority 6 mg, 12 mg, and 36 mg orforglipron to placebo will be conducted each at a 2-sided significance level of 0.05 using a 2-sample t-test. Additionally, a difference of at least 0.7 in mean change from baseline in PFS overall score at Week 72 for 6 mg, 12 mg, 36 mg, and pooled doses of orforglipron compared with placebo, and a common SD of 0.66 (Ullrich 2013).

6.9. Appendix 9: Statistical Analysis for China

General Considerations

Analyses will be performed for the following subpopulations

- participants enrolled in China
- participants enrolled in East Asian Countries/regions (mainland China, Japan, Taiwan, and Korea).

The analysis methods for the above-mentioned subgroups will be similar to those described for the main part of SAP GZGP. If there is not sufficient number of participants in the subpopulation, summary statistics will be provided.

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

6.10. Appendix 10: Statistical Analysis for Japan

Analyses will be performed for the following subpopulations

- participants enrolled in Japan
- participants who meet the Japan Society for the Study of Obesity (JASSO) subpopulation criteria.

The subpopulation analyses are subsets of the analyses described in the main part of SAP GZGP. The analyses to be included will be documented in a list of analyses which should include disposition, demographics, and selected efficacy and safety endpoints. If there is not a sufficient number of participants in this subpopulation, summary statistics will be provided.

JASSO subpopulation criteria

The JASSO subpopulation analysis will be performed according to the criteria of both BMI- and obesity-related health disorders according to the obesity disease label in Japan.

The target population is defined as participants who are diagnosed as having either hypertension, or dyslipidemia with:

1) baseline BMI between 27 to <35 kg/m^2, with at least 2 obesity-related health disorders, and
2) baseline BMI between ≥35 kg/m^2,

where the obesity-related health disorders are defined according to obesity disease treatment guideline (JASSO 2022). There are 11 obesity-related health disorders below Table GZGP.6.6. All participants and participants with obesity disease according to the JASSO treatment guideline will be compared.

Eleven obesity-related health disorders

The JASSO treatment guideline (JASSO 2022) defines 11 health disorders for the diagnosis of "obesity disease" in participants who need weight reduction for medical reasons.

Table GZGP.6.6. The 11 Obesity-Related Health Disorders

Obesity-related health disorders	Medical History Term
1) Glucose intolerance (including T2D, IGT)	Glucose tolerance impaired (IGT, etc.)
2) Dyslipidemia	Dyslipidemia
3) Hypertension	Hypertension
4) Hyperuricemia or Gout	Hyperuricaemia Gout
5) Cardiovascular disease, myocardial infarction, and angina	Coronary artery disease
6) Cerebral infarction and TIA	Cerebral infarction Transient ischemic attack
7) MASLD	Metabolic dysfunction-associated steatotic liver disease
8) Menstruation abnormalities and female infertility	Dysmenorrhoea Infertility Menstrual cycle abnormal
9) OSA and obesity-hypoventilation syndrome	Obstructive sleep apnoea
10) Motor dysfunction / Osteoarthritis	Motor dysfunction Osteoarthritis
11) Obesity-related renal disease	Renal disease

Abbreviations: IGT = impaired glucose tolerance; MASLD = Metabolic dysfunction-associated steatotic liver disease; OSA = obstructive sleep apnea; T2D = Type 2 diabetes; TIA = transient ischemic attack.

7. References

Andersen SW, Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013;12(1):7-16. https://doi.org/10.1002/pst.1548

Alosh, M, Bretz, F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med.* 2014;33(4):693-713. https://doi.org/10.1002/sim.5974

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

Cappelleri JC, Bushmakin AG, Gerber RA, et al. Evaluating the Power of Food Scale in obese subjects and a general sample of individuals: development and measurement properties. *Int J Obes (Lond)*. 2009;33(8):913-922. https://doi.org/10.1038/ijo.2009.107

[CDER/BIRRS] Center for Drug Evaluation and Research (CDER) and Biomedical Informatics and Regulatory Review Science (BIRRS) Team. Advancing premarket safety analytics workshop meet material_FDA-DM FMQs. Published September 06, 2022a. Accessed November 20, 2023. https://www.regulations.gov/document/FDA-2022-N-1961-0001

[CDER/BIRRS] Center for Drug Evaluation and Research (CDER) and Biomedical Informatics and Regulatory Review Science (BIRRS) Team. Standard Safety Tables and Figures: Integrated Guide. Published August 2022b. Accessed November 20, 2023. https://www.regulations.gov/document/FDA-2022-N-1961-0046

Chapman R. Expected a posteriori scoring in PROMIS®. *J Patient Rep Outcomes*. 2022;6(1):59. https://doi.org/10.1186/s41687-022-00464-9

Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Clarendon Press; 1994.

Du Y, Li J, Raha S, Qu Y. A unified Bayesian framework for bias adjustment in multiple comparisons from clinical trials. *Stat Med*. 2024;43(15):2928-2943. https://doi.org/10.1002/sim.10064

American Diabetes Association Professional Practice Committee. 6. Glycemic Goals and Hypoglycemia: Standards of Care in Diabetes-2025. *Diabetes Care*. 2025;48(Suppl1):S128-S145. https://doi.org/10.2337/dc25-s006

[EMA] European Medicines Agency. Guidance document for the content of the <Co->Rapporteur day 80 critical assessment report. EMA/269176/2014. Published 2014. Accessed May 3, 2023. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/day-80-assessment-report-clinical-guidance_en.pdf

EuroQol Research Foundation. EQ-5D-5L User Guide, Version 3.0. Updated September 2019. Accessed July 29, 2022. https://euroqol.org/publications/user-guides

[FDA] Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER). Adjusting for covariates in randomized clinical trials for drugs and biological products: Guidance for industry. Published May 2023. Accessed July 24, 2023.

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products

Flint A, Raben A, Blundell JE, Astrup A. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *Int J Obes Relat Metab Disord.* 2000;24(1):38-48. https://doi.org/10.1038/sj.ijo.0801083.

[ICH] International Council for Harmonisation. ICH harmonised guideline. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. E9(R1). Adopted on November 20, 2019. Accessed November 07, 2022. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

Inker LA, Schmid CH, Tighiouart H,et al.; CKD-EPI Investigators. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29. https://doi.org/10.1056/NEJMoa1114248.

[JASSO] Japan Society for the Study of Obesity. Guideline for the management of obesity disease [in Japanese]. 2022. Tokyo: Life Science Publishing. 2022.

Kolotkin RL, Ervin CM, Meincke HH, et al. Development of a clinical trials version of the Impact of Weight on Quality of Life-Lite questionnaire (IWQOL-Lite Clinical Trials Version): results from two qualitative studies. *Clin Obes*. 2017;7(5):290-299. https://doi.org/10.1111/cob.12197

Kolotkin RL, Williams VSL, Ervin CM, et al. Validation of a new measure of quality of life in obesity trials: Impact of Weight on Quality of Life-Lite Clinical Trials Version. *Clin Obes*. 2019;9(3):e12310. https://doi.org/10.1111/cob.12310

Lowe MR, Butryn ML, Didie ER, et al. The Power of Food Scale. A new measure of the psychological influence of the food environment. *Appetite* 2009;53(1):114-118. https://doi.org/10.1016/j.appet.2009.05.016

Ma C, Shen X, Qu Y, Du Y. Analysis of an incomplete binary outcome dichotomized from an underlying continuous variable in clinical trials. *Pharm Stat*. 2022;21(5):907-918. https://doi.org/10.1002/pst.2204

Maruish ME, editor. User's Manual for the SF36v2 Health Survey. 3rd ed. Lincoln, RI: Quality Metric Incorporated; 2011

PHUSE. Analysis and displays associated with adverse events: focus on adverse events in Phase 2-4 clinical trials and integrated summary documents. Published February 03, 2017. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+and+Displays+Associated+with+Adverse+Events+Focus+on+Adverse+Events+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf

PHUSE. Analysis and displays associated with demographics, disposition, and medications. Published March 02, 2018. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Demographics,+Disposition+and+Medication+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary+Documents.pdf

PHUSE. Analyses and displays associated with laboratory analyte measurements in Phase 2-4 clinical trials and integrated submission documents – update to recommendations. Published August 19, 2022. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Safety+Analytics/WP068.pdf

PHUSE. Analyses and displays associated with measures of central tendency – focus on vital sign, electrocardiogram, and laboratory analyte measurements in Phase 2-4 clinical trials and integrated submission documents. Published October 10, 2013. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Measures+of+Central+Tendency-+Focus+on+Vital+Sign,+Electrocardiogram+%26+Laboratory+Analyte+Measurements+in+Phase+2-4+Clinical+Trials+and+Integrated+Submissions.pdf

PHUSE. Analyses and displays associated with outliers or shifts from normal to abnormal: focus on vital signs, electrocardiogram, and laboratory analyte measurements in Phase 2-4 clinical trials and integrated summary documents. Published September 10, 2015. Accessed May 30, 2023. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Outliers+or+Shifts+from+Normal+To+Abnormal+Focus+on+Vital+Signes+%26+Electrocardiogram+%26+Laboratory+Analyte+Measurements+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf

Röver C, Bender R, Dias S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res Synth Methods*. 202112(4):448-474. https://doi.org/10.1002/jrsm.1475.

Sadoul BC, Schuring EAH, Mela DJ, Peters HPF. The relationship between appetite scores and subsequent energy intake: an analysis based on 23 randomized controlled studies. *Appetite*. 2014;83:153-159. https://doi.org/10.1016/j.appet.2014.08.016

Ullrich J, Ernst B, Wilms B, et al. Roux-en Y gastric bypass surgery reduces hedonic hunger and improves dietary habits in severely obese subjects. *Obes Surg*. 2013:23(1):50-55. https://doi.org/10.1007/s11695-012-0754-5

Wang B, Du Y. Improving the mixed model for repeated measures to robustly increase precision in randomized trials. *Int J Biostat*. 2023. https://doi.org/10.1515/ijb-2022-0101

Ye T, Bannick M, Yi Y, Shao J. Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. 2023. arXiv preprint. https://doi.org/10.48550/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

van Can J, Sloth B, Jensen CB, et al. Effects of the once-daily GLP-1 analog liraglutide on gastric emptying, glycemic parameters, appetite and energy metabolism in obese, non-diabetic adults. *Int J Obes (Lond)*. 2014;38(6):784-793. https://doi.org/10.1038/ijo.2013.162

Signature Page for VV-CLIN-117409 v3.0

Approval	PPD 25-Jun-2025 20:29:05 GMT+0000

Signature Page for VV-CLIN-117409 v3.0