Statistical Analysis Plan J2A-MC-GZGV (Version 3)

A Phase 3, Randomized, Open-Label Study to Investigate the Efficacy and Safety of Once Daily Oral Orforglipron Compared with Dapagliflozin in Adult Participants with Type 2 Diabetes and Inadequate Glycemic Control with Metformin (ACHIEVE-2)

NCT06192108

Approval Date: 04-Aug-2025

Title Page

Protocol Title: A Phase 3, Randomized, Open-Label Study to Investigate the Efficacy and Safety of Once Daily Oral Orforglipron Compared with Dapagliflozin in Adult Participants with Type 2 Diabetes and Inadequate Glycemic Control with Metformin (ACHIEVE-2)

Protocol Number: J2A-MC-GZGV

Compound Number: Orforglipron (LY3502970)

Short Title: Effect of Orforglipron Compared with Dapagliflozin in Participants with Type 2 Diabetes Inadequately Controlled with Metformin (ACHIEVE-2)

Acronym: ACHIEVE-2

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana, USA 46285

Regulatory Agency Identifier Number(s)

Registry	ID
IND	142842
EU trial number	2023-507206-13-00
WHO	U1111-1296-1507

Confidential Information

The information contained in this document is confidential and the information contained within it may not be reproduced or otherwise disseminated without the approval of Eli Lilly and Company or its subsidiaries.
Note to Regulatory Authorities: This document may contain protected personal data and/or commercially confidential information exempt from public disclosure. Eli Lilly and Company requests consultation regarding release/redaction prior to any public release. In the United States, this document is subject to Freedom of Information Act (FOIA) Exemption 4 and may not be reproduced or otherwise disseminated without the written approval of Eli Lilly and Company or its subsidiaries.

Document ID: VV-CLIN-134060

Approved on 04 Aug 2025 GMT

Table of Contents

Title Page 1
Table of Contents 2
Version history 6
1. Introduction 10
1.1. Objectives, Endpoints, and Estimands 10
1.2. Study Design 13
2. Statistical Hypotheses 16
2.1. Multiplicity Adjustment 16
3. Analysis Sets 18
4. Statistical Analyses 19
4.1. General Considerations 19
4.1.1. Baseline Definition 20
4.1.2. Analysis Methods 21
4.2. Participant Dispositions 30
4.3. Primary Endpoints/Estimands Analysis 31
4.3.1. Definition of Endpoint(s) 31
4.3.2. Main Analytical Approach 31
4.3.3. Sensitivity Analysis 31
4.4. Secondary Endpoints/Estimands Analysis 31
4.4.1. Key Secondary Endpoints/Estimands 31
4.4.2. Other Secondary Endpoints/Estimands 32
4.5. Tertiary and Exploratory Endpoints Analysis 36
4.5.1. Pharmacokinetic and Pharmacokinetic/Pharmacodynamic Methods 36
4.5.2. Exploratory Endpoints Analysis 36
4.6. Safety Analyses 36
4.6.1. Extent of Exposure 37
4.6.2. Adverse Events 38
4.6.3. Safety Topics of Interest 39
4.6.4. Additional Safety Assessments 47
4.7. Other Analyses 48
4.7.1. Subgroup Analyses 48
4.8. Interim Analyses 49
4.8.1. Data Monitoring Committee 49
4.9. Changes to Protocol-planned Analyses 49
5. Sample Size Determination 51

Approved on 04 Aug 2025 GMT

6. Supporting Documentation....52
6.1. Appendix 1: Demographic and Baseline Characteristics....52
6.2. Appendix 2: Historical Illnesses and Preexisting Conditions....53
6.3. Appendix 3: Treatment Compliance....54
6.4. Appendix 4: Prior/Concomitant Medications....54
6.5. Appendix 5: Important Protocol Deviations....54
6.6. Appendix 6: Statistical Analysis for China....54
6.7. Appendix 7: Empirical Estimation of Relative Event Rate....55
7. References....57

Approved on 04 Aug 2025 GMT

Table of Contents

Tables **Page**

Table GZGV.1.1. Visit Frequency 14
Table GZGV.4.1. Baseline and Postbaseline Definitions for Safety Analyses 21
Table GZGV.4.2. Estimands and Analysis Models for All Endpoints 23
Table GZGV.4.3. Imputation Procedure 26
Table GZGV.4.4. Study Discontinuation Reasons 27
Table GZGV.4.5. Listings and Summary Tables Related to Dispositions 31
Table GZGV.4.6. Listings and Summary Tables Related to AEs 38
Table GZGV.4.7. Description and Analyses of Safety Topics of Interest 40
Table GZGV.4.8. Summary Tables Related to Clinical Laboratory Evaluations 47
Table GZGV.4.9. Summary Tables Related to Vital Signs 47
Table GZGV.4.10. Summary Tables Related to ECG Parameters 48
Table GZGV.6.1. Demographics and Baseline Characteristics 53

Table of Contents

Figure **Page**

CCI [redacted] 14

CCI [redacted] 17

Figure GZGV.4.1. Treatment journey of a trial participant in relation to permanent discontinuation of study intervention or study....25

Version history

This Statistical Analysis Plan (SAP) for Study J2A-MC-GZGV (GZGV) is based on the GZGV Harmonized Clinical Protocol dated 17 August 2023.

The first amendment of SAP GZGV was updated based on regulatory feedback on the original SAP and protocol. The major updates in this amendment are listed in the table below.

SAP Version History Summary

SAP Version	Approval Date	Change section	Change	Rationale
1	15-Dec-2023		Not applicable	Original version
2	5-Nov-2024	1.1	Addition of text to the definition of population for treatment regimen and efficacy estimand.	Clarification
		2.1	[redacted]	CCI [redacted]
		4.1	Added clarification for different types of time-to-event analyses.	Clarification
		4.1.2.1.3	Further clarified the definition of Scenario 5. Removed the direct imputation method Added more details regarding the imputation methods Added footnote in Table GZGV.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 and Qu (2024) Clarified the distinction of imputing discontinuation due to protocol deviation, inadvertent enrolment
		4.1.2.3	Updated MMRM model for safety analyses, clarified the definition of the start time for different analyses.	Simplify the analysis model for safety analyses to be consistent with the PSAP.

Approved on 04 Aug 2025 GMT

		4.3.1	Added note that for missing or cancelled HbA1c measurement at endpoint visit, retests may be used.	Minimize the missing HbA1c in the primary endpoint analysis.
		4.4.2	Added definition of valid SMBG profile	Clarification.
		4.5	Updated Table GZGV.4.5.	To be consistent with other studies of the program.
		4.5.2	Remove exploratory analysis for HOMA.	This study does not collect insulin and C-peptide.
		4.6.3	Modified Table GZGV.4.7.	To be consistent with PSAP.
		4.7.1	Updated the testing of subgroup effects. Added details for the Bayesian shrinkage method.	Clarification.
		7	Reference section was updated.	To add extra references.
3	See date on page 1	1.1	Move [redacted] mg weight management objective from key secondary to secondary.	To be consistent with Protocol GZGV amendment b and graphical approach.
		4.1	Updated wording on which analyses will adjust for strata.	Clarification
		4.1.2	Clarified that risk difference will be used as the primary approach for treatment comparison for binary endpoints.	Clarification
		4.1.2.1.2	Updated logistic regression section to include interaction terms in model.	Consistency with other analyses.

Approved on 04 Aug 2025 GMT

		4.1.2.1.3	Further clarified the imputation strategy for all scenarios. Updated study discontinuation reason categories for imputation purpose.	Clarification
		4.1.2.3	Replaced negative binomial model with empirical method for analysis of hypoglycemia Clarified time-to-event analyses are based on unstratified methods	The approach aligns with the limited number of events
		4.6.2	Clarified sorting order of maximum severity CCI table. Added table to summarize dose reductions due to AE.	Clarification
		4.6.3	Removed redundant CCI NVDC by Maximum Severity table. Added shift tables for hepatic parameters.	Clarification
		4.7.1	Further clarified subgroup analysis section.	Clarification
		6.4	Updated time period of analysis for initiation of additional antihyperglycemic medication to “during treatment period.” Update the definition of baseline concomitant medication.	Clarification

		Appendix 8	Added Appendix 8.	Add statistical details for empirical estimation method.
		Throughout SAP	Minor changes and reorganization	For clarification, no change to analysis methodologies, so not detailed.
		Throughout SAP	Changed LS means to model-based estimated means	Update language to better align with unconditional average treatment effect estimate

Abbreviation: CCI ANCOVA = analysis of covariance; HbA1c = hemoglobin A1c, HOMA = homeostatic model assessment; MMRM = mixed model repeated measure, NVD = nausea, vomiting, diarrhea; NVDC = nausea, vomiting, diarrhea, constipation; PSAP = program safety analysis plan; SAP = statistical analysis plan; SMBG= self-monitoring of blood glucose; CCI

1. Introduction

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

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
To demonstrate that orforglipron [redacted] mg, [redacted] mg [redacted] mg are noninferior to dapagliflozin in glycemic control, [redacted]	• Change from baseline in HbA1c at Week 40
Key Secondary	
To demonstrate that orforglipron [redacted] mg, [redacted] mg, [redacted] mg are superior to dapagliflozin in glycemic control	• Change from baseline in HbA1c at Week 40
To demonstrate that orforglipron [redacted] mg, [redacted] mg, [redacted] mg are superior to dapagliflozin in additional measures of glycemic control	• Proportion of participants who achieved HbA1c <7.0% (53 mmol/mol) at Week 40 • Proportion of participants who achieved HbA1c ≤6.5% (48 mmol/mol) at Week 40
To demonstrate that orforglipron [redacted] mg [redacted] mg are superior to dapagliflozin in weight management	• Percentage change from baseline in body weight at Week 40 • Change from baseline in body weight at Week 40
To demonstrate that orforglipron [redacted] mg is superior to dapagliflozin in lipid parameters	• Percentage change from baseline in non-HDL cholesterol at Week 40 • Percentage change from baseline in triglycerides at Week 40
To demonstrate that orforglipron [redacted] mg is superior to dapagliflozin in SBP	• Change from baseline in SBP at Week 40
Other Secondary	
To demonstrate that orforglipron [redacted] mg is superior to dapagliflozin in weight management	• Percentage change from baseline in body weight at Week 40 • Change from baseline in body weight at Week 40
To compare orforglipron [redacted] mg, [redacted] mg, and [redacted] mg with dapagliflozin in additional measures of glycemic control	• Proportion of participants who achieved HbA1c [redacted] at Week 40 • Change from baseline in fasting serum glucose [redacted] at Week 40 [redacted]
To compare orforglipron [redacted] mg, [redacted] mg, and [redacted] mg with dapagliflozin in additional measurements of weight management	• Proportion of participants who achieved weight loss of ≥5%, ≥10%, and ≥15% from baseline at Week 40 [redacted]

Objectives	Endpoints
To compare orforglipron [redacted] mg and [redacted] mg with dapagliflozin in SBP	• Change from baseline in SBP at Week 40
To compare orforglipron [redacted] mg, [redacted] mg, and [redacted] mg with dapagliflozin in DBP	• Change from baseline in DBP at Week 40
To compare orforglipron [redacted] mg, [redacted] mg, and [redacted] mg with dapagliflozin in lipid parameters	• Percentage change from baseline in lipid parameters at Week 40 ○ total cholesterol [redacted]
To compare orforglipron [redacted] mg and [redacted] mg with dapagliflozin in lipid parameters	• Percentage change from baseline in lipid parameters at Week 40 ○ non-HDL cholesterol ○ triglycerides
To compare orforglipron [redacted] mg, [redacted] mg, [redacted] mg with dapagliflozin in patient-reported outcomes	• Change from baseline at Week 40 in ○ SF-36v2 Acute Form domain and summary scores [redacted] • DTSQc scores at Week 40

[redacted]

Abbreviations: [redacted] BMI = body mass index; DBP = diastolic blood pressure; DTSQc = Diabetes Treatment Satisfaction Questionnaire-Change; [redacted] HbA1c = hemoglobin A1c; HDL = high-density lipoprotein; [redacted] SBP = systolic blood pressure; SF-36v2, Acute = Short Form-36 version 2 Health Survey acute form; [redacted]

Primary estimands

There will be 2 estimands for the primary objective planned in the study. Details on plan for use of each estimand can be found in Section 2.1. The 2 estimands, treatment regimen estimand and efficacy estimand, address intercurrent events (ICEs) using the treatment policy strategy and the hypothetical strategy (FDA 2021), respectively.

Treatment regimen estimand

The clinical question of interest is: *What is the treatment difference in the change in HbA1c from baseline at Week 40 between orforglipron [redacted] mg compared with dapagliflozin [redacted] mg, orforglipron [redacted] mg compared with dapagliflozin [redacted] mg and orforglipron [redacted] mg compared with dapagliflozin [redacted] mg in individuals who meet eligibility criteria regardless of adherence to study intervention or initiation of additional antihyperglycemic medications?*

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.

Rationale for the treatment regimen estimand

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

Treatment regimen estimand attributes

- *Population*: Individuals who meet the eligibility criteria (Section 5.1and 5.2 of the Protocol). 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.
- *Endpoints*: Change from baseline in hemoglobin A1c (HbA1c) at Week 40
- *Treatment condition*: The randomized treatment with allowance for potential dose interruptions and dose modifications regardless of adherence to study intervention or initiation of additional antihyperglycemic medications
- *ICEs*: No ICEs are defined because treatment adherence and the initiation of additional antihyperglycemic medications are a part of the treatment condition.
- *Population-level summary and treatment effect of interest*: Difference in mean changes in HbA1c from baseline at Week 40:
    - orforglipron [redacted] mg compared with dapagliflozin [redacted] mg
    - orforglipron [redacted] mg compared with dapagliflozin [redacted] mg, and
    - orforglipron [redacted] mg compared with dapagliflozin [redacted] mg.

Efficacy estimand

The clinical question of interest is: *What is the treatment difference in the change in HbA1c from baseline at Week 40 between orforglipron [redacted] mg compared with dapagliflozin [redacted] mg, orforglipron [redacted] mg compared with dapagliflozin [redacted] mg and orforglipron [redacted] mg compared with dapagliflozin [redacted] mg in individuals who meet the eligibility criteria if they would remain on their randomly assigned study intervention for 40 weeks and would not initiate additional antihyperglycemic medications?*

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.

Rationale for the efficacy 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 antihyperglycemic medications.

Efficacy estimand attributes

- *Population*: Individuals who meet the eligibility criteria (Section 5.1 and 5.2 of the Protocol). 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.
- *Endpoints*: Change from baseline in HbA1c at Week 40
- *Treatment condition*: The randomized treatment with allowance for potential dose interruptions and dose modifications
- *ICEs*: ICEs include permanent discontinuation of study intervention and initiation of additional antihyperglycemic medications and will be 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 40 weeks and would not initiate additional antihyperglycemic medications. Dose modification and interruption will not be considered an ICE because they are part of the treatment condition.
- *Population-level summary and treatment effect of interest*: Difference in mean changes in HbA1c from baseline at Week 40:
    - orforglipron [redacted] mg compared with dapagliflozin [redacted] mg
    - orforglipron [redacted] mg compared with dapagliflozin [redacted] mg, and
    - orforglipron [redacted] mg compared with dapagliflozin [redacted] mg.

The same strategies will also be applied to the estimands for the key secondary endpoints as specified in Section 4.

1.2. Study Design

ACHIEVE-2 (Study J2A-MC-GZGV [GZGV]) is a Phase 3, multicenter, randomized, open-label (orforglipron dose blinded), active comparator-controlled, parallel-arm study designed to examine the efficacy and safety of orforglipron administered once daily (QD) compared with dapagliflozin, in participants with type 2 diabetes (T2D) with inadequate glycemic control using metformin (at least 1500 mg/day). The orforglipron doses ([redacted] mg, [redacted] mg, and [redacted] mg) will be double-blinded. Other details about blinding will be specified in the blinding and unblinding plan.

This study will consist of 3 periods CCI

- a screening and lead-in period: up to CCI
- a treatment period: 40 weeks, including CCI maintenance dose, and
- a safety follow-up period: CCI



CCI

The visit frequency is listed in Table GZGV.1.1.

Table GZGV.1.1. Visit Frequency

	Visit Number
Screening and lead-in period	601[a]
	1
	2
Treatment period	3
	4
	5
	6
	7
	8
	9
	9T[a]
	10

CCI

	Visit Number
	10T[a]
	11
	ED
Safety follow-up period	801

Abbreviation: ED = early discontinuation; T = telephone visit.

[a] Only for participants from China sites (Section 6.6).

Participants will be randomly assigned in a [redacted] ratio to orforglipron [redacted] mg, orforglipron [redacted] mg, orforglipron [redacted] mg, and dapagliflozin [redacted] mg, respectively.

The randomization will be stratified by country and [redacted].

2. Statistical Hypotheses

The numbering of the hypothesis below corresponds to objective, orforglipron dose, and indicator of null hypothesis.

The null hypotheses corresponding to the primary objective are as follows:

CCI

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

CCI

2.1. Multiplicity Adjustment

Multiplicity adjusted analyses will be performed on the primary and key secondary objectives in order 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 et al. (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).

CCI illustrates the current graphical testing procedure. Since more than one types of estimands are used to assess primary and key secondary objectives and are intended for separate and independent purposes, no multiplicity adjustment will be made for conducting separate analyses on the same objectives. Treatment regimen estimand will be the primary estimand, with the efficacy estimand considered supportive. There will be no adjustment for multiple comparisons for any other analyses beyond the primary and key secondary objectives.



CCI

3. Analysis Sets

For the purposes of analysis, the following analysis sets are defined:

Participant Analysis Set	Description
Entered participants	All participants who sign informed consent[a]
Randomized participants	All participants who are randomly assigned a study intervention
Safety participants	All participants who are randomly assigned a study intervention and who take at least 1 dose of study intervention

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

The following data point sets are defined:

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 additional antihyperglycemic medications[a]
Safety data points set	All data points obtained during the treatment period and the follow-up period defined as at or after baseline and up to the date of study withdrawal/study completion including the safety follow-up period regardless of study intervention discontinuation or initiation of additional antihyperglycemic medications[a]
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 additional antihyperglycemic medications[a]

[a] "Additional antihyperglycemic medications" refers to any antihyperglycemic therapy that is used for more than 2 weeks (14 days).

Approved on 04 Aug 2025 GMT

4. Statistical Analyses

4.1. General Considerations

Statistical analysis will be the responsibility of Eli Lilly and Company (Lilly) or its designee. Some analyses and summaries described in this analysis plan may not be conducted if warranted by data (that is, too few events to justify conducting an analysis). Additional analyses of the data may be conducted as deemed appropriate without further changes made to the protocol or statistical analysis plan (SAP), even after the final database lock.

Unless otherwise noted, all tests of treatment effects will be conducted at the 2-sided α level of 0.05, and the confidence intervals (CIs) will be calculated at 95%, 2-sided.

Participants will be analyzed according to the treatment group to which they are randomly assigned.

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. Some specific predefined parameters may be log-transformed before statistical analysis, if deemed necessary. If 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 continuous measures, summary statistics will include sample size, mean, standard deviation (SD), median, minimum, and maximum for the actual, the change from baseline, and if applicable, the percent change from baseline measurements as well. Model-based estimated means and standard errors derived from the analysis models will also be displayed for the actual value, change or percent change from baseline measurements. For analysis of log-transformed parameters, model estimated means and standard errors on the original scale will be derived through back-transformation using the delta method from the model-based estimated means and standard errors on the natural log-scale. Treatment comparisons will be displayed showing the model-based estimated means, the 95% CIs, and the p-values for the treatment differences. All baseline measures will be analyzed using an analysis of variance (ANOVA) model that has the treatment group as the fixed effect.

For categorical measures, summary statistics will include sample size, frequency, and percentages. For safety categorical measures, risk difference and 95% CI will be provided wherever applicable. If applicable, Fisher's exact test or Pearson's chi-squared test will be used for treatment comparisons.

For time-to-event measures, participants without an event will be censored, and time-to-event will be the number of days between the start date and the date of the participant's end of follow-up + 1 day (depending on the estimand definition). For participants experiencing the event, "time-to-first-event" will be the time (in days) from start date to the first occurrence of the event + 1 day. For safety and study intervention discontinuation analyses, the date of first dose of study intervention will be used as the start date. For other analyses (e.g. study discontinuation), the randomization date will be used as the start date (in rare cases where randomization occurred prior to visit 3 date, Visit 3 date will be used instead).

Approved on 04 Aug 2025 GMT

Data may exist at postbaseline visits where a variable is not scheduled to be collected. Data from unscheduled visits and from early discontinuation (ED) visits that do not correspond to a scheduled visit will be excluded from mixed-model for repeated measures (MMRM) analysis, analysis of covariance (ANCOVA), or logistic regression analysis, unless otherwise specified.

For stratification factors, countries with fewer than 10 randomized participants will be pooled into 1 category (pooled country). Unless otherwise specified, country and baseline HbA1c level (≤8.0%, >8.0%) will be used in the analysis models including MMRM, ANCOVA, and logistic regression analysis.

No data collected at the optional prescreening visit (V601), except adverse events (AEs) and demographics information, will be included in statistical analyses, given that the purpose prescreening visit is not for data collection (Protocol GZGV, Section 1.3) but for operational efficiency.

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.

A subset of the planned efficacy, health outcomes, and safety analyses will be reproduced based on participants enrolled in China and East Asian countries. See Appendix 6 (Section 6.6) for details.

4.1.1. Baseline Definition

Unless otherwise specified, baseline for efficacy assessments is defined as the last available nonmissing measurement prior to the first dosing of study intervention, which in most cases will be the measurement recorded at Week 0 (Visit 3). If there are no doses of study intervention, baseline will be defined as the last available nonmissing measure on or prior to randomization. For patient-reported outcome measures, data obtained at Visit 3, regardless of the timing relative to first dose, will serve as baseline. For the safety related parameters, the definition of baseline and postbaseline are specified in Table GZGV.4.1.

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

Analysis Type	Participant Analysis Set/Data Points Set	Baseline	Postbaseline
Treatment-emergent adverse events	Safety participants/safety data points set	Start of screening[a] and ends prior to the first dose of study intervention (typically at Week 0)	Starts after the first dose of study intervention and ends at the end of the safety follow-up period or the date of study withdrawal [b].
Treatment-emergent abnormal laboratory values, vital signs, and ECGs	Safety participants/safety data points set	Start of screening[a] 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 safety follow-up period 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	Safety participants/safety data points set	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 safety follow-up period or the date of study withdrawal. Only scheduled and early termination visits will be included.

Abbreviation: ECG = electrocardiogram.

[a] Adverse data collected between Visit 601 and Visit 1 will be included as part of screening as well.

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

4.1.2. Analysis Methods

The primary and key secondary objectives, as well as blood pressure, lipid parameters, and patient-reported parameters, will be evaluated based on both treatment regimen estimand and efficacy estimand. All other secondary endpoints (except summary of safety data) will be guided by the efficacy estimand. The assessments guided by the treatment regimen estimand will be conducted using the randomized participants analysis set and the treatment regimen estimand data points set. The assessments guided by the efficacy estimand will be conducted using the randomized participants analysis set and the efficacy estimand data points set (Section 3).

Unless otherwise specified, safety and tolerability assessments will be guided by an estimand comparing safety of orforglipron treatment groups with the dapagliflozin treatment group for the entire study period (the treatment period plus safety follow-up), irrespective of adherence to study intervention or initiation of additional antihyperglycemic medications for all study population. It is noted that for some safety endpoints, such as treatment discontinuation due to AEs and laboratories measured that are supposed to be measured only during the treatment period, the analyses will be based on data observed during the treatment period only.

Approved on 04 Aug 2025 GMT

The estimands, participant analysis sets, analysis data points sets, and analysis models for efficacy and safety endpoints are summarized in Table GZGV.4.2.

Table GZGV.4.2. Estimands and Analysis Models for All Endpoints

	Treatment Regimen Estimand	Efficacy Estimand	Estimand Guiding Safety
Participants	Randomized	Randomized	Safety
Data points set	Treatment regimen	Efficacy estimand	Safety
Primary and key secondary endpoints **Selected other secondary endpoints** • blood pressure • lipid parameters • patient-reported outcomes	X	X	
Other secondary endpoints • additional measures of glycemic control • additional measures of weight management		X	
Safety (not including lipid parameters)			X
Analysis model	• Continuous: ANCOVA with missing endpoints imputed[a] • Binary[c]: logistic regression with missing endpoints imputed[a]	• Continuous: MMRM (if there are multiple longitudinal postbaseline measurements) or ANCOVA with missing endpoints imputed[b] (if there is only 1 postbaseline measurement) • Binary[c]: logistic regression with missing endpoints imputed[b]	• Continuous: MMRM (if there are multiple longitudinal postbaseline measurements) or ANCOVA with no imputation to be conducted (if there is only 1 postbaseline measurement) • Binary: risk difference (95% CI); Fisher's exact or Pearson's chi-squared test if applicable; no imputation will be conducted • Time to event: Cox Regression and Kaplan-Meier.

Abbreviations: ANCOVA = analysis of covariance; CI = confidence interval; MMRM = mixed model repeated measures.

[a] The treatment regimen estimand imputation method is described in Section 4.1.2.1.3.

[b] The efficacy estimand imputation method is described in Section 4.1.2.2.3.

Approved on 04 Aug 2025 GMT

[c] The statistical testing for risk difference will be used as the primary approach for treatment comparison.

4.1.2.1. Analysis Methods for Treatment Regimen Estimand

4.1.2.1.1. Method for Continuous Variables

For the assessment of continuous efficacy measures guided by the treatment regimen estimand, data for participants with missing values at Week 40 will be imputed based on methods described in Section 4.1.2.1.3 and then analyzed using an ANCOVA with a variance estimator that is robust to model misspecification and heteroscedasticity (Ye et al. 2022, FDA 2023). The model will include

- treatment group as a factor variable
- pooled country as a factor variable
- baseline HbA1c level CCI as a factor variable
- baseline value (of the dependent variable)
- interaction between treatment group and baseline HbA1c level CCI and
- interaction between treatment group and baseline value.

Statistical comparisons will only be performed between each dose of orforglipron and dapagliflozin 10 mg. If the model fails to converge, the interaction terms will be removed. 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 from using Rubin's Rule, which combines estimates from multiple imputed datasets. The imputation procedure for creating a single imputed dataset detailed in Table GZGV.4.3.

Some parameters may be log-transformed before fitting the ANCOVA as specified in Sections 4.4 through 4.5.2, with the associated log-transformed baseline value as a covariate. In these cases, model-based estimated means and 95% CIs for each treatment group and treatment difference will be back-transformed and presented as mean percent change from baseline and as relative treatment difference to dapagliflozin 10 mg in percent change.

4.1.2.1.2. Method for Binary Variables

For binary efficacy measures derived from a continuous variable and guided by the treatment regimen estimand, the missing value in the underlying continuous variable will be imputed first based on methods described in Section 4.1.2.1.3 and then the corresponding binary variable will be derived.

For the assessment of binary efficacy measures, a logistic regression model will be used to examine the treatment difference with missing values at Week 40 imputed based on methods described in Section 4.1.2.1.3. The model will include:

- treatment group as a factor variable
- pooled country as a factor variable
- baseline HbA1c level CCI as a factor variable
- baseline value (of the dependent variable)
- interaction between treatment group and baseline HbA1c level CCI and
- interaction between treatment group and baseline value.

The unconditional treatment group effect will be assessed by risk difference and relative risk based on the delta-method using formula provided (Ye et al. 2023). The final inference will be derived using Rubin's Rule, which combines estimates from multiple imputed datasets, with the imputation procedure for creating a single imputed dataset detailed in Table GZGV.4.3. The estimated treatment group-specific risk, risk difference, relative risk, p-value, and 95% CI will be presented.

4.1.2.1.3. Primary Multiple Imputation (PMI) Strategy

Discontinuation is expected to be uncommon. Participants who discontinue study intervention (that is, discontinue study treatment) will be encouraged to continue in the study for treatment period/phase and follow-up period/phase (note, the words period and phase are used interchangeably in this SAP, while the case report forms [CRFs] use "phase"). 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 will be imputed in a manner 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 40 visit) shown in Figure GZGV.4.1.

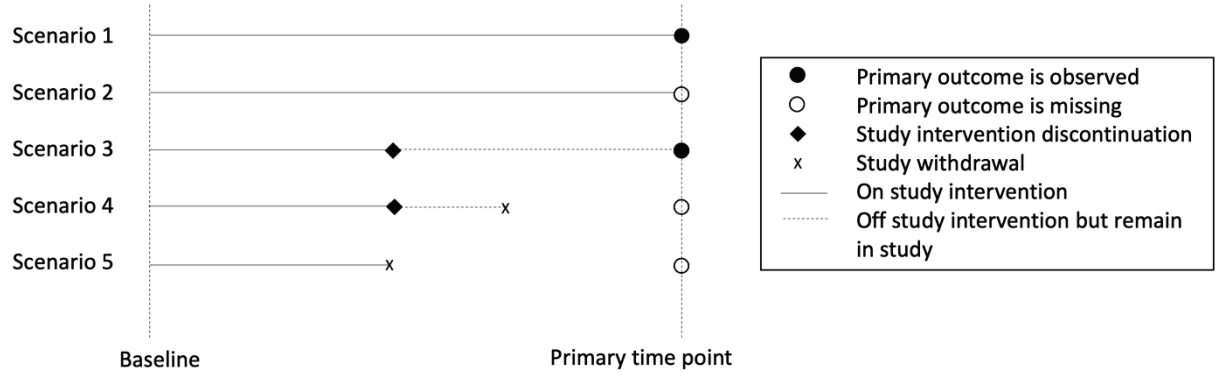


Figure GZGV.4.1. Treatment journey of a trial participant in relation to permanent discontinuation of study intervention or study.

Scenario 5 also 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.

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 study intervention but continued in Study GZGV 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, model implemented does not converge), an alternative multiple imputation method (return to baseline approach) will be used. Any missing values at baseline will be imputed in the same model when imputing the endpoint.

Table GZGV.4.3. Imputation Procedure

Scenario (Study Intervention/Study Discontinuation; Missingness 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 post-baseline visits, across all time points from Scenario 1 and Scenario 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 study discontinuation at a later visit or completion of treatment period; with missing value	Missing values at the endpoint visit will be imputed by treatment group through MCMC (predictive mean matching method), using baseline and endpoint visit data from Scenario 3 (MI-RD). 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 study discontinuation reasons classified as Category 5A in Table GZGV.4.4), 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 study discontinuation reasons classified as Category 5B in Table GZGV.4.4), 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-RD = multiple imputation using retrieved dropouts; N/A = not applicable.

Table GZGV.4.4. Study Discontinuation Reasons

Disposition Reason	Associated Subcategories	Category
Adverse event		5A
Death		5A
Protocol deviation		5A
Pregnancy		5A
Non-compliance with study drug		5A
Lack of efficacy		5A
Withdrawal by subject	Concern about study procedures/perceived risks	5A
	Scheduling conflicts	5A
	Subject is moving or has moved	5A
	Personal issue unrelated to trial	5A
	Due to epidemic/pandemic	5B
	Other (option to include a specify field)	5A
	Concern about study procedures/perceived risks	5A
Physician decision	Due to epidemic/pandemic	5B
	Other (option to include a specify field)	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, e.g. Adverse Event (5A).

4.1.2.1.4. Modified Multiple Imputation Strategy

A modified multiple imputation 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.

4.1.2.1.5. 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 HbA1c change post treatment period discontinuation by different treatment groups could impact the treatment comparisons.

To start with, missing data are imputed according to the imputation procedure as shown in Table GZGV.4.3. A penalty is then added to those imputed values at Week 40 visit. The analysis is intended to vary the magnitude of the penalties added for all treatment groups under comparison and evaluate the impact these would have on the study conclusion. A 2-dimentional space of penalties will be assessed for each pair comparison of orforglipron and dapagliflozin.

MI-TP aims to evaluate the robustness of the noninferiority claim to the assumptions of using the observed data to impute the missing HbA1c in all treatment groups.

4.1.2.2. Analysis Methods for Efficacy Estimand

4.1.2.2.1. Method for Continuous Variables

For the assessment of continuous efficacy measures guided by the efficacy estimand, a restricted maximum likelihood (REML)-based MMRM (Wang and Du 2024) analysis will be used. All the longitudinal observations at each scheduled postbaseline visit will be included in the analysis. The MMRM model will include the following terms:

- treatment group as a factor variable
- visit as a factor variable
- pooled country as a factor variable
- baseline HbA1c level CCI [redacted] as a factor variable
- baseline value (of the dependent variable)
- interaction between treatment group, visit, and baseline HbA1c level CCI [redacted] and
- interaction between treatment group, visit, and baseline value.

The estimated treatment group effect and comparison between each of the orforglipron treatment group and the dapagliflozin treatment group at the scheduled visits will be reported together with the variability estimated using the robust inference as provided in Wang and Du (2023). The associated 2-sided 95% CI and corresponding p-values will also be reported. The sandwich estimator (Diggle et al. 1994) for the variance-covariance matrix will be used. An unstructured covariance matrix by treatment groups will be used to model the within-participant errors, assuming heteroscedasticity (variant covariance structure across treatment groups) and the measurements for different participants are independent. If the model fails to converge, the model is 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
- pooled country as a factor variable
- baseline HbA1c level CCI [redacted] as a factor variable
- baseline value (of the dependent variable)
- interaction between visit and treatment group
- interaction between visit and baseline HbA1c level CCI [redacted] and
- interaction between visit and baseline value

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

- unstructured covariance matrix
- heterogeneous Toeplitz
- heterogeneous autoregressive
- heterogeneous compound symmetry

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

The first covariance structure that converges will be used.

Some parameters may be log-transformed before fitting the MMRM as specified in Sections 4.4 through 4.5.2. In these cases, model-based estimated means and 95% CIs for each treatment group and treatment difference will be back-transformed and presented as mean percent change from baseline and as relative treatment difference to the dapagliflozin treatment group in percent change.

If the data does not warrant the MMRM model, then an ANCOVA model will be conducted as described in Section 4.1.2.1.1. Missing values will be imputed based on methods described in Section 4.1.2.2.3.

4.1.2.2.2. Method for Binary Variables

For binary efficacy measures derived from a continuous variable and guided by the efficacy estimand, the missing value in the underlying continuous variable will be imputed first based on methods described in Section 4.1.2.2.3, and then the corresponding binary variable will be derived.

For the assessment of binary efficacy objectives, the logistic regression model described in Section 4.1.2.1.2 will be used after imputation and estimation of treatment group specific risk, risk difference, relative risk, p-value, and 95% CI will be presented.

4.1.2.2.3. Handling of Missing Data

Missing data should be minimized at the best precaution. For continuous efficacy analyses guided by efficacy estimand, the hypothetical strategy will be used to handle the 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. For ANCOVA analysis, missing values at the last visit 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 assumption. The same method will also be used to impute the missing value in the underlying continuous variables of binary outcomes.

4.1.2.3. Analysis Methods for Safety

For categorical safety measures, treatment group differences of percentages will be conducted using Fisher's exact test or Pearson's chi-squared test if applicable, unless otherwise specified. Risk difference and the 95% CI will be provided.

For selected continuous safety measures, unless otherwise specified, treatment group differences of mean change or mean percent change from baseline at all scheduled visits will be assessed via an MMRM using restricted maximum likelihood REML, 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 fixed effects 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, p-amylase, and lipase) may be log-transformed before fitting the MMRM as specified in Section 4.5.2. In these cases, model-based estimated means and 95% CIs 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 the dapagliflozin treatment group in percent change.

For selected safety parameters, time-to-first-event analysis via unstratified Cox proportional hazards model may be conducted. P-values will be calculated based on an unstratified log-rank test. The Kaplan-Meier method will be used to estimate the cumulative event curve over time. Counts and proportions of participants who experience the event will be calculated by treatment group.

Where necessary, the rate of events will be analyzed using an empirical method (see Appendix 7) for statistical comparison.

Some safety analysis may be conducted after excluding data after the initiation of new antihyperglycemic medications.

4.2. Participant Dispositions

The participant dispositions for the screening period, the study intervention/treatment, the treatment period, and/or the follow-up period will be collected in CRFs with the corresponding primary reason. The study completion for a participant is defined as the participant completing both the treatment period and the follow-up period, regardless of completion of study treatment.

The planned listings and summary tables for dispositions are proved in Table GZGV.4.5.

Table GZGV.4.5. Listings and Summary Tables Related to Dispositions

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

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

4.3. Primary Endpoints/Estimands Analysis

The primary study objective is to demonstrate that orforglipron CCI mg, CCI mg, CCI mg are noninferior to dapagliflozin in glycemic control, CCI

4.3.1. Definition of Endpoint(s)

The primary efficacy measure is change from baseline in HbA1c at Week 40. The change from baseline in HbA1c for each participant at each nominal visit is defined as postbaseline HbA1c at the nominal visit minus baseline, where baseline is defined in Section 4.1.

In addition, if HbA1c is missing/canceled at the primary endpoint visit (Week 40), retests performed within 7 days of Week 40 will be used as primary endpoint in model analyses.

4.3.2. Main Analytical Approach

The analytical approaches are specified for the treatment regimen estimand in Section 4.1.2.1.1 (ANCOVA) and the efficacy estimand in Section 4.1.2.2.1 (MMRM).

For each hypothesis, orforglipron treatment groups will be declared noninferior to dapagliflozin in controlling HbA1c if the upper limit of the CI interval for the model-based estimated mean difference in the change from baseline in HbA1c for orforglipron minus dapagliflozin is below 0.3%, with the CI calculated at the α level allocated to the hypothesis according to the graphical approach
(Section 2.1).

4.3.3. Sensitivity Analysis

The sensitivity analyses are specified in Section 4.1.2.1.4 and Section 4.1.2.1.5 (MI-TP).

4.4. Secondary Endpoints/Estimands Analysis

4.4.1. Key Secondary Endpoints/Estimands

The key secondary study objectives are listed in Section 1.1.

4.4.1.1. Definition of Endpoints

The definition of the endpoints are

- change from baseline in HbA1c at Week 40
- incidence of participant (proportion at the population-level) achieving HbA1c target value of <7.0% (53 mmol/mol) at Week 40
- incidence of participant (proportion at the population-level) achieving an HbA1c target value of ≤6.5% (48 mmol/mol) at Week 40
- percentage change from baseline in body weight at Week 40 (CCI mg and CCI mg only)
- change from baseline in body weight at Week 40 (CCI mg and CCI mg only)
- percentage change from baseline in non-HDL cholesterol at Week 40 (CCI mg only)
- percentage change from baseline in triglycerides at Week 40 (CCI mg only), and
- change from baseline in SBP at Week 40 (CCI mg only),

4.4.1.2. Main Analytical Approach

The analytical approaches are specified for the treatment regimen estimand in Section 4.1.2.1 and the efficacy estimand in Section 4.1.2.2.

For each hypothesis, orforglipron treatment groups will be declared superior to dapagliflozin if the p-value is less than the α level allocated to the hypothesis according to the graphical approach (Section 2.1).

4.4.2. Other Secondary Endpoints/Estimands

The following other secondary efficacy endpoints will be guided by the efficacy estimand (Section 4.1.2.2).

- incidence of participant (proportion at the population-level) achieving a HbA1c target value of CCI at Week 40
- change from baseline in fasting serum glucose at Week 40

CCI

- incidence of participant (proportion at the population-level) achieving a weight loss of ≥5%, ≥10%, and ≥15% from baseline at Week 40

CCI

For SMBG, valid SMBG profiles will be used for analysis, defined as having non-missing values at ≥ time points (including at least one post-meal point) among the 7 pre-specified time points and being collected within 2 weeks prior to the given visit. For each time point, the average of the corresponding SMBG values from the valid SMBG profiles will be used for analysis. In addition to the analyses on the mean of all 7-point measurements, similar analyses will be done for each of the 7 points, the 2-hour morning, midday, and evening meal excursions, the mean of all meals' 2-hour excursion, the mean of all premeal measurements, and the mean of all 2-hour postprandial measurements.

The following other secondary efficacy endpoints will be guided by the treatment regimen estimand (Section 4.1.2.1) and the efficacy estimand (Section 4.1.2.2).

- percentage change from baseline in body weight at Week 40 ([redacted] mg only)
- change from baseline in body weight at Week 40 ([redacted] mg only)
- change from baseline at Week 40 in
  - SBP ([redacted] mg and [redacted] mg only), and
  - diastolic blood pressure
- percentage change from baseline in lipid parameters at Week 40 (log transformed data)
  - total cholesterol

[redacted]

  - non-HDL cholesterol ([redacted] mg and [redacted] mg only), and
  - triglycerides ([redacted] mg and [redacted] mg only)
- change from baseline at Week 40 in
  - Short Form-36 version 2 Health Survey (SF-36v2) Acute Form domain and summary scores

[redacted]

- score of the Diabetes Treatment Satisfaction Questionnaire-Change (DTSQc) at Week 40.

The analytical approaches for the other secondary efficacy endpoints are specified for the treatment regimen estimand in Section 4.1.2.1 and the efficacy estimand in Section 4.1.2.2, unless otherwise stated. The analytical approaches for the other secondary safety endpoints are specified in Section 4.1.2.3.

No multiplicity adjustment will be conducted for non–key secondary endpoints.

4.4.2.1. Health Outcomes

The patient-reported outcome questionnaires will be completed by the participants at baseline and postbaseline (Week 40 or ED visit). Guided by the treatment-regimen estimand and efficacy estimand, the analysis will be conducted using an ANCOVA model as described in Section 4.1.2.1.1, with missing values at Week 40 being imputed based on the methods described in Section 4.1.2.1.3 and Section 4.1.2.2.3, respectively for the 2 estimands.

4.4.2.1.1. Short Form-36 Health Survey Version 2, Acute Form

Per copyright owner, the Procore® 2.0 Smart Measurement System will be used to derive the following 8 domains and 2 component scores:

- physical functioning domain
- role-physical domain
- bodily pain domain
- general health domain
- vitality domain
- social functioning domain
- role-emotional domain
- mental health domain
- mental component score, and
- physical component score.

Each domain is scored individually, and information from these 8 domains is further aggregated into the 2 health component scores. Items are answered on Likert scales of varying lengths (3-point, 5-point, or 6-point scales). Scoring of each domain and both component scores are norm-based and presented in the form of T-scores, with a mean of 50 and a SD of 10; higher scores indicate better levels of function and/or better health (Maruish 2011). The following analyses for the actual values and change from baseline values for the 8 domains and 2 component scores will be conducted at Week 40:

- descriptive summaries by treatment group, and
- ANCOVA analysis as described in Section 4.4.2.1.

4.4.2.1.2. EQ-5D-5L

The EQ-5D-5L is a standardized 5-item instrument that assesses 5 dimensions of health:

- mobility
- self-care
- usual activities
- pain/discomfort, and
- anxiety/depression.

Each dimension has 5 levels:

- no problems
- slight problems
- moderate problems
- severe problems, and
- extreme problems.

The EQ-5D-5L also has a vertical VAS, called the EQ VAS, which records the respondent's self-rated health, with endpoints labeled as "best imaginable health state" (100) and "worst imaginable health state" (0). EQ-5D-5L health states are converted into a single index 'utility' score using a scoring algorithm based on public preferences. The single index 'utility' score will be based on the appropriate United Kingdom scoring algorithm (Devlin 2018), where higher scores indicate better overall health related quality of life.

Each item will be summarized descriptively by treatment group at each scheduled visit at which the EQ-5D-5L is administered. The changes from baseline to Week 40 in the index and VAS scores will be analyzed using an ANCOVA as described in Section 4.4.2.1.

4.4.2.1.3. Ability to Perform Physical Activities of Daily Living

The APPADL contains 7 items that assess how difficult it is for participants to engage in certain activities considered to be integral to normal daily life, such as walking, standing, and climbing stairs (Hayes et al. 2012). Items are scored on a 5-point numeric rating scale, where 5 = "not at all difficult" and 1 = "unable to do." A raw overall score is calculated by simply summing the scores of the 7 items, and a transformed overall score is obtained by linearly transforming the raw overall score to a 0 to 100 scale. A higher raw overall score and a higher transformed overall score are indicative of better ability to perform activities of daily living.

Descriptive summaries by treatment group at each scheduled visit at which the APPADL is administered will be presented for each item. Treatment comparison in the transformed overall APPADL score change from baseline to Week 40 will be analyzed using an ANCOVA as described in Section 4.4.2.1.

4.4.2.1.4. Impact of Weight on Self-perceptions Questionnaire

The IW-SP contains 3 items that assess how often a participant's body weight affects how happy they are with their appearance and how often they feel self-conscious when out in public (Hayes and DeLozier 2015). Each item is rated on a 5-point scale ranging from 1 = "always" to 5 = "never." Total scores for the IW-SP are derived by summing the item scores and dividing by the number of items. The score can also be transformed to a range from 0 to 100. Higher IW-SP scores correspond to better self-perception.

Descriptive summaries by treatment group at each scheduled visit at which the IW-SP is administered will be presented for each item. Treatment comparison in the transformed overall IW-SP total score change from baseline to Week 40 will be analyzed using an ANCOVA as described in Section 4.4.2.1.

4.4.2.1.5. Impact of Weight on Quality of Life-Lite Clinical Trials Version

The IWQOL-Lite-CT (Kolotkin et al. 2017, 2019) is a 20-item, obesity-specific, patient-reported outcomes instrument developed for use in weight management clinical studies.

The IWQOL-Lite-CT assesses 2 primary domains of obesity-related, health-related quality of life: Physical (7 items) and Psychosocial (13 items). A 5-item subset of the Physical domain – the Physical Function composite is also supported. 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 3 domain scores and total score range from 0 to 100 with higher scores indicating greater functioning.

Descriptive summaries by treatment group at each scheduled visit at which the IWQOL-Lite-CT is administered will be presented for each item. Treatment comparison in the raw and transformed overall IWQOL-Lite-CT total score change from baseline to Week 40 will be analyzed using an ANCOVA as described in Section 4.4.2.1.

4.4.2.1.6. Diabetes Treatment Satisfaction Questionnaire

The Diabetes Treatment Satisfaction Questionnaire (DTSQ) contains 8 items (conceptually the same items in the DTSQs and DTSQc versions) (Bradley 1994).

- For DTSQs, 6 items (1 and 4 through 8) are rated from 0 to 6, and item scores are summed to produce a measure of treatment satisfaction. Greater scores indicate greater treatment satisfaction. The 2 remaining items (2 and 3) are treated individually to assess, respectively, the perceived frequency of hyperglycemia and hypoglycemia.
- For DTSQc, 6 items (1 and 4 through 8) are scored on a scale of −3 to +3. Item scores are summed to produce a measure of change in treatment satisfaction. For all items except Item 2 (perceived frequency of hyperglycemia) and Item 3 (perceived frequency of hypoglycemia)
    - the higher the score, the greater the improvement in treatment satisfaction
    - the lower the score, the greater the deterioration in treatment satisfaction
    - a score of 0 represents no change, and
    - for Items 2 and 3, the lower the score, the better the glucose control.

Descriptive summaries will be provided at baseline (DTSQs only) and at Week 40 (DTSQc only) for the perceived hyperglycemia item, perceived hypoglycemia item, and the 6-item overall satisfaction score.

Treatment comparison in the change from baseline in DTSQs scores and in the actual responses to DTSQc at Week 40 will be analyzed using an ANCOVA as described in Section 4.4.2.1, except that baseline will not be used as a covariate in DTSQc. The analyses will be conducted for the perceived hyperglycemia item, perceived hypoglycemia item, and the 6-item overall satisfaction score.

4.5. Tertiary and Exploratory Endpoints Analysis

4.5.1. Pharmacokinetic and Pharmacokinetic/Pharmacodynamic Methods

Not applicable.

4.5.2. Exploratory Endpoints Analysis

Additional exploratory analysis may be conducted as deemed necessary.

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, 2015, 2017, 2018, 2022) and publications from regulatory agencies (for example, FDA 2010; EMA 2014; CDER/BIRRS 2022, 2023). Descriptions of the safety analyses are provided in this SAP, but some details are found in compound level safety standards.

Unless otherwise specified, safety analyses will be conducted using the safety participants and the safety data points set (Table GZGV.4.1). It is noted that for some safety endpoints, such as, study intervention discontinuation due to AE, and so on, the analyses will be based on data

observed during treatment period only. The analytical approaches for safety analyses are specified in Section 4.1.2.3. Summary tables with risk difference will be sorted by decreasing order in risk difference. Not all displays will necessarily be created as a "static" display. Some will be incorporated into interactive display tools instead of or in addition to a static display.

4.6.1. Extent of Exposure

Duration of exposure to study intervention will be summarized by treatment group for safety participants. Exposure will be summarized and calculated for the treatment period being considered as the date of last dose of study intervention minus the date of first dose of study intervention plus 1 day using safety participants. Duration on study (date of end of study participation – date of randomization + 1) will also be summarized by treatment group.

Descriptive statistics (including n, mean, SD, minimum, first quartile, median, third quartile, and maximum) will be provided, and the frequency of participants falling into different exposure ranges will be summarized. Overall exposure will be summarized in total patient-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 and percentages of participants falling into the following range will be summarized by planned treatment group as well:

- >0
- ≥4 weeks
- ≥8 weeks
- ≥12 weeks
- ≥16 weeks
- ≥20 weeks
- ≥24 weeks
- ≥32 weeks, and
- ≥40 weeks.

In addition, the frequency and percentages of participants falling into the following exposure ranges for study and study intervention may be summarized by planned treatment group:

- 0 weeks
- >0 to <4 weeks
- ≥4 to <8 weeks
- ≥8 to <12 weeks
- ≥12 to <16 weeks
- ≥16 to <20 weeks
- ≥20 to <24 weeks
- ≥24 to <32 weeks
- ≥32 to <40 weeks, and
- ≥40 weeks

Participants with dose interruptions/modifications will be summarized with reasons by treatment group.

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

4.6.2. Adverse Events

The planned summaries and listings for AEs are provided in Table GZGV.4.6, and details are described in compound level safety standards.

Table GZGV.4.6. Listings and Summary Tables Related to AEs

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 intervention discontinuation due to an AE)
TEAEs by PT within SOC	Fisher's exact	Safety/TP + FP
Maximum Severity TEAEs		Safety/TP + FP
TEAEs with incidence ≥5% 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
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 of AEs of suspected overdosing of orforglipron		Safety/TP
Listing of participants with at least 1 notable event		Randomized/TP + FP

Abbreviations: CCI eCRF = electronic case report form; FP = follow-up period; PT = preferred term; CCI SOC = system organ class; CCI TP = treatment period.

4.6.3. 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 1 or more standardized Medical Dictionary for Regulatory Activities (MedDRA) query(ies) (SMQs), system organ class (SOC), high-level term, FDA Medical Query, 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. Searching criteria are detailed in the compound level safety standards.

The planned analyses for safety topics of interest are provided in Table GZGV.4.7 and are described more fully in compound level safety standards.

Table GZGV.4.7. Description and Analyses of Safety Topics of Interest

Safety Topic of Interest	Short Description	Analysis	Method	Population/Period
MACE	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 and PT (if necessary) MACE reported by investigators may be summarized	Fisher's exact	Safety/TP + FP
		Listing of all MACE reported by investigator (whether or not positively adjudicated) (if necessary)		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.	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 AE database will be searched using predefined PTs.	TE hypotension, orthostatic hypotension and syncope disorders by PT[a]	Fisher's exact	Safety/TP + FP
Hypoglycemia		Incidence (percent of participants experiencing 1 or more episode) and the rate (episodes/participant/year) of Level 3 (severe/serious) hypoglycemia	Logistic regression (for incidence) and empirical method (for rate)	
		Incidence and rate of Level 2 or Level 3 hypoglycemia		

Approved on 04 Aug 2025 GMT

Safety Topic of Interest	Short Description	Analysis	Method	Population/Period
	The 2023 American Diabetes Association position statement on glycemic targets (ElSayed et al. 2023) will be used to define: • Level 1 hypoglycemia • Level 2 hypoglycemia • Level 3 hypoglycemia (severe/serious hypoglycemia)	Incidence and rate of Level 1 hypoglycemia (if warranted by data)		Primary analysis: Safety/TP + FP excluding events after initiation of new antihyperglycemic therapy (regardless of how many days of use) Supportive analysis: Safety/TP + FP regardless of initiation of new antihyperglycemic therapy
		Listing of Level 2 or Level 3 hypoglycemia events		Safety/TP + FP
Severe persistent hyperglycemia	Initiation of new antihyperglycemic therapy taken for severe persistent hyperglycemia	Rescue therapy taken for severe persistent hyperglycemia (if data warrants)[a]	Fisher's exact	Safety/TP + FP
GI adverse events	GI AEs using gastrointestinal disorders SOC will be captured.	Severe or serious TE gastrointestinal events by PT	Fisher's exact	Safety/TP + FP
		Plots of prevalence and incidence over time for TE nausea, vomiting, diarrhea, constipation, and nausea, vomiting, and diarrhea (NVD) by maximum severity		Safety/TP + FP
		Summary of prevalence and incidence over time for TE nausea, vomiting, diarrhea, constipation, and NVD		Safety/TP + FP
		Plot of time to the onset of TE nausea, vomiting, diarrhea, constipation, and NVD	Kaplan-Meier	Safety/TP + FP
		Study intervention discontinuation due to GI	Fisher's exact	Safety/TP + FP
Renal safety		Shift of minimum-to-minimum for eGFR estimated by the CKD-EPI equation[b]		Safety/TP + FP

Safety Topic of Interest	Short Description	Analysis	Method	Population/Period
	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 identified using SMQs also.	Shift of maximum-to-maximum for UACR		Safety/TP + FP
		MMRM analyses for eGFR[b]	MMRM	Safety/TP + FP
		MMRM analyses for UACR (log transformation)	MMRM	Safety/TP + FP
		TE renal events by PT nested within SMQs[a]	Fisher's exact	Safety/TP + FP
		TE dehydration events by PT[a]	Fisher's exact	Safety/TP + FP
Pancreatitis	The pancreatic enzyme data (p-amylase and lipase) will be observed through laboratory testing.	Shift of maximum-to-maximum for pancreatic enzyme		Safety/TP + FP
		MMRM analysis for pancreatic enzymes: p-amylase and lipase (log transformation)	MMRM	Safety/TP + FP
		Investigator-reported events and positively adjudicated pancreatic events, respectively	Fisher's exact	Safety/TP + FP
		TE Pancreatic events by "Acute pancreatitis" SMQ and "Chronic pancreatitis" PT	Fisher's exact	Safety/TP + FP
		Listing of adjudicated and investigator-reported pancreatitis		Safety/TP + FP

Safety Topic of Interest	Short Description	Analysis	Method	Population/Period
	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.			
Thyroid malignancies and C-cell hyperplasia	TE thyroid malignancies and C-cell hyperplasia will be identified using MedDRA HLTs and PTs. The purpose of calcitonin measurements is to assess the potential effect of orforglipron on thyroid C-cell function.	TE thyroid C-cell hyperplasia and malignancies by PT	Fisher's exact	Safety/TP + FP
		Shift of maximum-to-maximum for calcitonin value in the thresholds		Safety/TP + FP
		MMRM analysis for calcitonin (log transformation)	MMRM	Safety/TP + FP
		Postbaseline calcitonin value of ≥35 ng/mL that has increased at least 50% over baseline[a]	Fisher's exact	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

Approved on 04 Aug 2025 GMT

Safety Topic of Interest	Short Description	Analysis	Method	Population/Period
Hepatic safety	Hepatic laboratory values include ALT, AST, serum ALP, TBL, DBL, INR, 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
		TE potentially drug-related hepatic disorders by PT nested within SMQ	Fisher's exact	Safety/TP + FP
		Hepatocellular drug-induced liver injury screening plot (TBL vs ALT or AST)		Safety/TP + FP
		Hepatocellular drug-induced liver injury screening table		Safety/TP + FP
		Listing of participants with ALT or AST≥ 3X ULN		Safety/TP + FP
		Cholestatic drug-induced liver injury screening plot (TBL vs ALP)		Safety/TP + FP
		Cholestatic drug-induced liver injury screening table		Safety/TP + FP
		Listing of participants with ALP or TBL ≥2X ULN		Safety/TP + FP
		Participant profiles for participants meeting criteria for a comprehensive hepatic evaluation (as defined in the protocol)		Safety/TP + FP
		Shift of maximum-to-maximum for ALT, AST, ALP, TBL		Safety/TP + FP

Approved on 04 Aug 2025 GMT

Safety Topic of Interest	Short Description	Analysis	Method	Population/Period
Gallbladder and biliary tract disorders	All events of TEAE biliary colic, cholecystitis, or other suspected events related to gallbladder disease 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	AEs will be searched using MedDRA PTs from narrow scope FMQ or SMQs.	TE allergic reactions and hypersensitivities by PT nested within applicable FMQ or SMQ[a]	Fisher's exact	Safety/TP + FP
Major depression, suicidal ideation, and behavior	AEs will be searched using MedDRA PTs that satisfies the search criteria.	TE depression, suicidal ideation, and behavior events by PT nested within applicable FMQ or SMQ[a]	Fisher's exact	Safety/TP + FP
Abuse potential	AEs will be searched using a modified abuse potential FMQ.	TE abuse potential events by PT[a]	Fisher's exact	Safety/TP + FP
Diabetic retinopathy	Diabetic retinopathy and macular edema will be monitored using retinal fundus photography and additional ophthalmological follow-up as	Baseline and postbaseline diabetic retinopathy assessment by treatment group and by eye (that is, any eye, left eye, and right eye)		Safety/TP + FP
		TE diabetic retinopathy, diabetic macular edema, and related complications by PTs	Fisher's exact	Safety/TP + FP
		Shift of baseline diabetic retinopathy and macular edema status to most severe postbaseline diabetic retinopathy and macular edema status		Safety/TP + FP

Safety Topic of Interest	Short Description	Analysis	Method	Population/Period
	warranted throughout the study for participants with T2D. Diabetic retinopathy complication will be searched using MedDRA PTs.			

Abbreviations: CCI ALT = alanine aminotransferase; ALP = alkaline phosphatase; AST = aspartate aminotransferase; CEC = clinical endpoint committee; CKD-EPI = Chronic Kidney Disease Epidemiology; CV = cardiovascular; DBL = direct bilirubin; eCRF = electronic case report form; eGFR = estimated glomerular filtration rate; FDA = United States Food and Drug Administration; FMQ = FDA Medical Query; FP = follow-up period; GGT = gamma-glutamyltransferase; GI = gastrointestinal; HLT = high-level term; INR = international normalized ratio; MACE = major adverse cardiovascular event; MedDRA = Medical Dictionary for Regulatory Activities; MMRM = mixed model repeated measure; NVD = nausea, vomiting, diarrhea; PT = preferred term; SMQ = standardized MedDRA query; SOC = system organ class; T2D = type 2 diabetes; TBL = total bilirubin; CCI TP = treatment period; UACR = urinary albumin to creatinine ratio; vs = versus.

[a] For these tables, if the number of events is less than 10, a listing will be provided instead.

[b] eGFR will be calculated by Chronic Kidney Disease-Epidemiology cystatin-c equation (Inker et al. 2012).

4.6.4. Additional Safety Assessments

Actual and change from baseline to postbaseline values of central laboratory measures, vital signs, and selected electrocardiogram (ECG) parameters will be summarized at each scheduled visit. Continuous variables, as well as the change from baseline for these variables, will be analyzed by MMRM or ANCOVA models as described in Section 4.1.2.3.

4.6.4.1. Central Laboratory Measures

All laboratory data will be reported in the International System of Units (SI). Selected laboratory measures will also be reported using conventional units.

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

Table GZGV.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 of 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.

4.6.4.2. Vital Signs

Measurements taken in triplicate for a particular visit will be averaged before any analysis or summary outlined in this section.

The planned summaries for vital signs (SBP, diastolic blood pressure, and pulse rate) are provided in Table GZGV.4.9, and are described more fully in compound level safety standards.

Table GZGV.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 pulse rate and change from baseline	MMRM	Safety/TP + FP
Summary of participants meeting specific blood pressure and pulse rate levels	Descriptive statistics	Safety/TP + FP
Shift of maximum-to-maximum for pulse rate		Safety/TP + FP

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

4.6.4.3. Electrocardiograms

Measurements taken in triplicate for a particular visit will be averaged before any analysis or summary outlined in this section. The planned summaries for ECG parameters are provided in Table GZGV.4.10 and are described more fully in compound level safety standards.

Table GZGV.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
Analysis for change from baseline in ECG parameters (heart rate, PR interval)	MMRM	Safety/TP + FP
Heart rate and PR interval in specified categories	Descriptive statistics	Safety/TP + FP

Abbreviations: CI = confidence intervals; ECG = electrocardiogram; FP = follow-up period; MMRM = mixed model repeated measure; PR = pulse rate; SD = standard deviation; TP = treatment period.

4.7. Other Analyses

4.7.1. Subgroup Analyses

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

- Change from baseline in HbA1c at Week 40 visit

The analyses will be conducted across the following subgroups (but not limited to):

- age group (<65, ≥65 years)
- age group (<75, ≥75 years)
- sex (female, male)
- baseline HbA1c stratum (≤8.0%, >8.0%)
- baseline BMI (<30, ≥30 and <35, ≥35 kg/m^2)
- baseline eGFR calculated by CKD-EPI cystatin-c equation (<60 and ≥60 mL/min/1.73 m^2)
- duration of T2D (<median, ≥median)
- duration of T2D (≤5, >5 and ≤10, >10 years)
- race
- ethnicity
- geographic region (Unites States versus outside of the United States), and
- country

The ANCOVA model specified in Section 4.1.2.1.1 will be fitted separately within each category of subgroup. The model-based estimated means, estimated treatment difference, SE, and 95% CI will be presented. The model-based estimated 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 any category within the subgroup (for example, status of yes or no) is <10% of the total population, only descriptive statistics will be provided for that category (that is, no inferential testing will be performed within the subgroup category).

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:

$$Y_i \sim N(\mu_i, \sigma_i^2),$$
$$\mu_i \sim N(\mu, \tau^2),$$
$$\mu \sim N(0, 5^2),$$
$$\tau \sim N^+(1),$$

where σ_i^2 is set to the observed variance for sample estimate, $N^+(1)$ is the half-normal distribution with a scaled parameter of 1. 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 2% based on historical T2D studies, and a standard deviation of 5% is chosen for the centrality parameter μ , so that its standard deviation is approximately 2.5 times the UISD. A scale parameter of 1 is chosen for the half-normal distribution such that the median heterogeneity in HbA1c reduction is approximately 0.67%, which is a reasonably large difference.

Additional subgroup evaluations may be conducted as exploratory analyses.

4.8. Interim Analyses

No interim analyses are planned for this study. If an unplanned interim analysis is deemed necessary for reasons other than a safety concern, the protocol must be amended.

4.8.1. Data Monitoring Committee

No Data Monitoring Committee (DMC) is 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 LY3502970 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 T2D. The DMC will have the responsibility to review unblinded interim safety analysis results from that study. The DMC may be asked to review unblinded safety data from Study GZGV 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

In Section 5, the power calculation assumptions were updated based on the treatment regimen estimand results in historical studies with the sample size remaining the same. To reflect this, the following changes are made in this document:

- Based on the SD of the modified intent-to treat population from Study I8F-MC-GPGL (SURPASS-2) (GPGL), SD was updated from 1.1% to 1.2%.

Approved on 04 Aug 2025 GMT

- "use of 2-sample t-test using HbA1c data" was further specified to be for the treatment regimen estimand.
- Sentence about dapagliflozin and orforglipron treated participant dropout rate was removed.

In addition, Section 5 has further clarification on the split of the alpha for the power calculation. Finally, in Section 5, a sentence with the name of the software used to calculate the sample size was added.

The objective to demonstrate that orforglipron 3 mg are superior to dapagliflozin in weight management was excluded from the graphical multiple testing procedure. Thus, in Section 2, there is no mention of the null hypothesis related to this objective and is not included in CCI

5. Sample Size Determination

Approximately 888 participants will be randomly assigned in a CCI ratio to orforglipron CCI mg, orforglipron CCI mg, orforglipron CCI mg, and dapagliflozin CCI mg, respectively.

This trial is powered at approximately 90% to assess noninferiority of orforglipron CCI mg to dapagliflozin CCI mg, orforglipron CCI mg noninferiority to dapagliflozin CCI mg, and/or orforglipron CCI mg noninferiority to dapagliflozin CCI mg, respectively, at the 2-sided CCI significance level (in the graphical approach, each of the 2 primary objectives related to the CCI mg and the CCI mg are tested at an α level of CCI and CCI respectively, and the primary objective related to the CCI mg is initially tested at an α level of CCI), relative to the primary endpoint (change from baseline in HbA1c at Week 40), under the following assumptions:

- use of a 2-sample t-test to compare treatment means using HbA1c data for the treatment regimen estimand
- CCI% greater reduction in HbA1c from baseline for CCI mg, CCI mg, and CCI mg of orforglipron compared to CCI mg of dapagliflozin

CCI

- common SD of CCI%.

Furthermore, this sample size will ensure approximately CCI% power for superiority of orforglipron CCI mg to dapagliflozin CCI mg, orforglipron CCI mg compared with dapagliflozin CCI mg, and orforglipron CCI mg compared with dapagliflozin CCI mg assuming CCI% greater HbA1c reduction of orforglipron CCI mg compared with dapagliflozin CCI mg, orforglipron CCI mg compared with dapagliflozin CCI mg, and orforglipron CCI mg compared with dapagliflozin CCI mg.

These assumptions are considered to ensure the power for the primary endpoint for both efficacy and treatment regimen estimands.

The power calculations were performed with the software nQuery version 9.1.0.0.

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics

Demographics and other baseline characteristics will be summarized by treatment group for randomized participants. The listing of basic demographic characteristics (that is, age, sex, ethnicity, race, country, baseline body weight, and so on) for randomized participants will be provided.

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

Table GZGV.6.1 describes the specific variables and how they will be summarized. Additional variables may also be included in the final study report.

Table GZGV.6.1. Demographics and Baseline Characteristics

Variable	Quantitative Summary	Categorical Summary
Age[a]	Yes	<65, ≥65 years
		<75, ≥75 years
		<65, ≥65 and <75, and ≥75 years
Sex	No	Male, Female
Ethnicity	No	Hispanic/Latino, Non-Hispanic/Non-Latino
Race	No	American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian or other Pacific Islander, White, or Multiple
Country/geographic region	No	By each country and pooled region
Height (cm)	Yes	
Baseline tobacco use	No	Never, Current, Former
Baseline waist circumference (cm)	Yes	
Baseline body weight (kg)	Yes	
Baseline BMI (kg/m^2)	Yes	<30, ≥30 and <35, ≥35 kg/m^2
Duration of T2D (years)	Yes	≤5, >5 and ≤10, >10 years
Baseline systolic blood pressure (mmHg)	Yes	
Baseline diastolic blood pressure (mmHg)	Yes	
Baseline pulse rate (beats/min)	Yes	
Baseline eGFR (CKD-EPI) (mL/min/1.73m^2)[b]	Yes	<45, ≥45 to <60, ≥60 to <90, ≥90 mL/min/1.73 m^2
Baseline UACR (g/kg)	Yes	<30, ≥30 and ≤300, >300 g/kg
Baseline HbA1c (%)	Yes	≤8.0%, >8.0%
Baseline HbA1c (mmol/mol)	Yes	
Baseline fasting serum glucose (mg/dL)	Yes	
Baseline fasting serum glucose (mmol/L)	Yes	
Baseline fasting lipid parameters (SI)	Yes	
Baseline fasting lipid parameters (CN)	Yes	

Abbreviations: BMI = body mass index; CKD-EPI = Chronic Kidney Disease Epidemiology; eCRF = electronic case report form; eGFR = estimated glomerular filtration rate; HbA1c = hemoglobin A1c; T2D = type 2 diabetes; UACR = urinary albumin to creatinine ratio.

[a] Age in years will be calculated as length of the time interval from the imputed date of birth (01 July in the year of birth collected in the eCRF) to the informed consent date.

[b] eGFR will be calculated by Chronic Kidney Disease-Epidemiology cystatin-c equation (Inker et al. 2012).

6.2. Appendix 2: Historical Illnesses and Preexisting Conditions

The count and percentages of participants with historical illnesses and preexisting conditions will be summarized by treatment group using the MedDRA PTs nested within SOC. The SOCs will be in alphabetical order. Events will be ordered by decreasing frequency. Conditions (that is, PTs) will be ordered by decreasing frequency within SOC. This will be summarized for all randomized participants. Historical illnesses are illnesses that end prior to informed consent, and preexisting conditions are conditions that are still ongoing at informed consent. No statistical comparisons between treatment groups will be performed.

6.3. Appendix 3: Treatment Compliance

Treatment compliance is defined as taking at least 75% of scheduled 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%*

If the data warrants, frequency counts and percentages of participants who are noncompliant to treatment and/or have dose interruptions will be summarized by treatment group for safety participants. Listings of such participants may also be provided.

6.4. Appendix 4: 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 first dose date of 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.

Concomitant medication will be summarized by PTs, by treatment group, and by decreasing frequency for randomized participants. Additionally, the following prior/concomitant medications of interest will be summarized by treatment group:

- prior glucagon-like peptide-1 use prior to study entry
- baseline antihypertensive therapy, by type
- baseline lipid lowering therapy, by type
- utilization of the following medication during treatment period
    - antihypertensive therapy
    - lipid lowering therapy
- initiation of additional antihyperglycemic medication during treatment period
- initiation of following medications during treatment period and safety follow-up period
    - antidiarrheal medication
    - antiemetic medication, and
    - constipation medication.

6.5. Appendix 5: Important Protocol Deviations

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

6.6. Appendix 6: Statistical Analysis for China

Analyses will be performed for the following subpopulations:

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

- participants enrolled in mainland China

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

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

6.7. Appendix 7: Empirical Estimation of Relative Event Rate

Traditionally, the Poisson distribution has been assumed to draw inference for the rate of rare events. When the event is rare and the sample size is large, it is known that the overall number of events is approximately from Poisson distribution. However, for some not very rare events such as severe hypoglycemic events in T2D patients, the total number of events may not be distributed from Poisson and may be over-dispersed. Assuming Poisson distribution may significantly underestimate the variance, and therefore may reduce the average probability and inflate the Type I error. An empirical method in estimating the variance of event rate without assuming any distribution on the number of events will be provided in this appendix.

Let X_{ij} denote the count response variable for subject *j* in treatment group *i*. Let $Y_i = \sum_j X_{ij}$ be the total number of events for treatment group *i*, and T_i denote the exposure for treatment group *i*. Let *i* = 0 for the control group and *i* = 1 for the experimental treatment group. The event rate for treatment group *i* can be calculated as

$$\hat{r}_i = \frac{Y_i}{T_i}$$

The empirical variance of $\hat{r}_i$ is

$$\widehat{Var}(\hat{r}_i) = T_i^{-2}\widehat{Var}(Y_i) = T_i^{-2} n_i S_i^2,$$

where S_i^2 is the variance of X_{ij} for treatment group *i*. Using the delta-method, the variance of $\log(\hat{r}_i)$ can be estimated as

$$\widehat{Var}(\log(\hat{r}_i)) = Y_i^{-2} n_i S_i^2$$

The relative rate of the experimental treatment versus the control treatment is estimated as

$$\hat{\lambda} = \frac{\hat{r}_1}{\hat{r}_0}$$

The variances of $\hat{\lambda}$ and $\log(\hat{\lambda})$ are

$$\widehat{Var}(\hat{\lambda}) = \hat{\lambda}^2 \widehat{Var}(\log(\hat{\lambda}))$$

$$\widehat{Var}\left(\log\left(\hat{\lambda}\right)\right) = \widehat{Var}(\log(\hat{r}_0)) + \widehat{Var}(\log(\hat{r}_1)) = Y_0^{-2} n_0 S_0^2 + Y_1^{-2} n_1 S_1^2$$

Assuming $\log(\hat{\lambda})$ is asymptotically from a normal distribution, the $100(1-\alpha)\%$ confidence interval for $\log(\hat{\lambda})$ can be constructed as

$$\left[\log\left(\hat{\lambda}\right) - z_{1-\frac{\alpha}{2}}\sqrt{\widehat{Var}\left(\log\left(\hat{\lambda}\right)\right)},\ \log\left(\hat{\lambda}\right) + z_{1-\frac{\alpha}{2}}\sqrt{\widehat{Var}\left(\log\left(\hat{\lambda}\right)\right)}\right]$$

Then, the $100(1-\alpha)\%$ confidence interval for $\hat{\lambda}$ is

$$\left[\hat{\lambda}\exp\left(-z_{1-\frac{\alpha}{2}}\sqrt{\widehat{Var}(\log(\hat{\lambda}))}\right),\quad \hat{\lambda}\exp\left(z_{1-\frac{\alpha}{2}}\sqrt{\widehat{Var}(\log(\hat{\lambda}))}\right)\right] \tag{1}$$

The p-value for testing the null hypothesis of $H_0: \lambda = 1$ is calculated as

$$p = 2\Phi\left(\left|\log(\hat{\lambda})\right| / \sqrt{\widehat{Var}(\log(\hat{\lambda}))}\right) \tag{2}$$

where Φ is cumulative distribution function of the standard normal distribution.

7. References

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

CCI [redacted]

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

[CDER/BIRRS] Center for Drug Evaluation and Research and Biomedical Informatics and Regulatory Review Science Team. Advancing premarket safety analytics workshop meet material FDA-DM FMQs. Published September 06, 2022. Accessed November 16, 2023. https://www.regulations.gov/document/FDA-2022-N-1961-0001

[CDER/BIRRS] Center for Drug Evaluation and Research and Biomedical Informatics and Regulatory Review Science Team. Standard safety tables and figures: integrated guide. Published February 08, 2023. Accessed November 16, 2023. https://www.regulations.gov/document/FDA-2022-N-1961-0046

Devlin NJ, Shah KK, Feng Y, et al. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ*. 2018;27(1):7-22https://doi.org/10.1002/hec.3564.

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

ElSayed NA, Aleppo G, Aroda VR, et al.; American Diabetes Association. 6. Glycemic targets: standards of care in diabetes-2023. *Diabetes Care*. 2023:46(suppl 1):S97-S110. https://doi.org/10.2337/dc23-S006

[EMA] European Medicines Agency. Guidance document for the content of the <co->rapporteur day 80 critical assessment report. Published 2014. Accessed 17 November 2022. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/day-80-assessment-report-clinical-guidance_en.pdf

[FDA] United States Food and Drug Administration. E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials: guidance for industry. Published May 2021. Accessed September 12, 2023. https://www.fda.gov/media/148473/download

[FDA] United States Food and Drug Administration. 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

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.

CCI [redacted]

CCI [redacted]

Liu R, Qu Y, Retrieved dropout imputation considering administrative study. 2024. https://arxiv.org/abs/2410.06774

Maruish ME, editor. *User's Manual for the SF-36v2 Health Survey*. 3rd ed. Quality Metric Inc.; 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 November 16, 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. Updated March 02, 2018. Accessed November 16, 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 June 16, 2022. Accessed November 16, 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 November 16, 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 November 16, 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*. 2021;12(4):448-474. https://doi.org/10.1002/jrsm.1475

Wang B, Du Y. Improving the mixed model for repeated measures to robustly increase precision in randomized trials. *Int Biostat J.* 2023;20(2):585-598. https://doi.org/10.1515/ijb-2022-0101

Wang Y, Tu W, Koh W, et al. Bayesian hierarchical models for subgroup analysis. *Pharm Stat*. 2024;23(6):1065-1083. https://doi.org/10.1002/pst.2424

Ye T, Shao J, Yi Y, Zhao Q. Toward better practice of covariate adjustment in analyzing randomized clinical trials. *J Am Stat Assoc*. 2022;118(544):2370-2382. https://doi.org/10.1080/01621459.2022.2049278

Ye T, Bannick M, Yi Y, Shao J. Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes. *Stat Theory Relat Fields*. 2023;7(2):159-163. https://doi.org/10.1080/24754269.2023.2205802

Signature Page for VV-CLIN-134060 v3.0

Approval	PPD Statistician 04-Aug-2025 02:00:31 GMT+0000

Signature Page for VV-CLIN-134060 v3.0