Statistical Analysis Plan Version 1 H9X-MC-GBGO

A Randomized, Double-Blind Trial Comparing the Effect of the Addition of Dulaglutide 1.5 mg versus the Addition of Placebo to Titrated Basal Insulin on Glycemic Control in Chinese Patients with Type 2 Diabetes (AWARD-CHN3: Assessment of Weekly AdministRation of LY2189265 in Diabetes-CHINA3)

NCT04591626

Approval Date: 29-Jun-2022

1. Statistical Analysis Plan: A Randomized, Double-Blind Trial Comparing the Effect of the Addition of Dulaglutide 1.5 mg versus the Addition of Placebo to Titrated Basal Insulin on Glycemic Control in Chinese Patients with Type 2 Diabetes (AWARD-CHN3: Assessment of Weekly AdministRation of LY2189265 in Diabetes-CHINA3)

Dulaglutide (LY2189265)

Study H9X-MC-GBGO is a Phase 3b, randomized, double-blind, placebo-controlled trial that compares the effect of the addition of dulaglutide 1.5 mg versus the addition of placebo to titrated basal insulin on glycemic control in Chinese patients with type 2 diabetes.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol H9X-MC-GBGO
Phase 3b

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly on date provide below.

Approved on 29 Jun 2022 GMT

2. Table of Contents

Table of Contents

Table **Page**

List of Figures

Figure **Page**

3. Revision History

The protocol for this study was approved on 06-March-2020. SAP Version 1.0 was approved prior to first patient visit.

SAP Version 2.0 was proved prior to first unblinding. Key changes include:

- Clarified classification of hypoglycemia events based on white paper Version 5.
- Incorporated details on multiple imputation using jump to reference as sensitivity analysis.
- Clarified shift tables for selected analytes.
- Removed analysis for pancreatic amylase.
- Added Standardized MedDRA Queries (SMQs) and Lilly Search Categories (LSC) for adverse events of special interest (AESI).
- Modified Treatment-emergent adverse events (TEAE) definition to be more precise.
- Removed analyses for hypoglycemia events with post-rescue or early-discontinuation data.
- Removed interaction term and covariate in negative binomial model for hypoglycemia events.
- Modified shift tables for selected analytes.
- Added baseline insulin dose in subgroup analysis.
- Updated number of patients required for subgroup analysis.
- Added Chinses EQ-5D-5L score in addition to UK score.

4. Study Objectives

Table 4.1Table 4.1 Objectives and Endpoints presents the objectives and endpoints of the study.

Table 4.1 Objectives and Endpoints

Objectives	Endpoints
Primary • To show the superiority of the addition of dulaglutide 1.5 mg QW compared to the addition of once weekly placebo to titrated insulin glargine, with metformin and/or acarbose, on change from baseline in HbA1c after 28 weeks of treatment in adult Chinese patients with T2D.	• The change in HbA1c from baseline to Week 28.
Key Secondary (controlled for Type I error) *Efficacy:* • To compare the addition of dulaglutide 1.5 mg QW with the addition of placebo QW to titrated basal insulin glargine for other parameters of glucose control and body weight change after 28 weeks of treatment in patients with T2D.	• Proportion of patients achieving HbA1c target of <7.0%. • Change from baseline in body weight. • Change from baseline in FSG (central lab).
Additional Secondary *Efficacy:* • To compare the addition of dulaglutide 1.5 mg QW with the addition of placebo QW to titrated basal insulin glargine for the following parameters after 28 weeks of treatment in patients with T2D.	• Proportion of patients achieving HbA1c target of ≤6.5%. • Proportion of patients achieving HbA1c target of <7.0% and with no weight gain (<0.1 kg) at 28 weeks, and without documented symptomatic hypoglycemia (BG <3.0 mmol/L) during the maintenance period (Weeks 12-28). • Proportion of patients achieving HbA1c target of <7.0% at 28 weeks, and without documented symptomatic hypoglycemia (BG <3.0 mmol/L) during the maintenance period (Weeks 12-28). • Proportion of patients achieving HbA1c target of <7.0%, and without weight gain (<0.1 kg) at 28 weeks. • Changes from baseline in BG from daily SMBG profiles. • Changes from baseline in daily mean insulin glargine doses.

Safety: • To compare the addition of dulaglutide 1.5 mg QW with the addition of placebo QW to titrated basal insulin glargine for selected safety parameters through 32 weeks in patients with T2D (unless noted otherwise).	• Incidence of TEAEs, SAEs and discontinuation of study drug due to AEs. • Change from baseline in electrocardiograms, laboratory and vital sign measurements. • Incidence of AESI. • The incidence of deaths and nonfatal CV events. • Occurrence of hypoglycemic episodes. • Incidence of patients initiating rescue therapies due to severe, persistent hyperglycemia. • Incidence of adjudicated and confirmed pancreatitis. • Incidence of thyroid neoplasm. • Incidence of allergic and hypersensitivity AEs.
Exploratory *Efficacy:* • To compare the addition of dulaglutide 1.5 mg QW with the addition of placebo QW to titrated basal insulin glargine for exploratory efficacy parameters after 28 weeks of treatment in patients with T2D.	• Frequency of insulin dose adjustments through 28 weeks. • Number of insulin dose assessments per TTT algorithm and compliance with the TTT algorithm at Week 6, 8, and 12. • Proportion of patients achieving HbA1c target of <7.0%, and without weight gain (<0.1 kg) at Week 28, and without documented symptomatic hypoglycemia (BG <3.9 mmol/L) during the maintenance period (Weeks 12-28). • Proportion of patients achieving HbA1c target <7.0% and without weight gain (<0.1 kg) at Week 28; • Proportion of patients achieving HbA1c target of <7.0% at Week 28, and without documented symptomatic hypoglycemia (BG <3.9 mmol/L) during the maintenance period.
Safety: • To compare the addition of dulaglutide 1.5 mg QW with the addition of placebo QW to titrated basal insulin glargine for exploratory safety parameters through 32 weeks in patients with T2D.	• Assessments of time to onset and duration of diarrhea, nausea, abdominal distension and vomiting in dulaglutide-treated patients.
Health Outcome/Quality of Life Measures: • To compare the addition of dulaglutide 1.5 mg QW with the addition of placebo QW to titrated basal insulin glargine for the following HO/QoL parameters in patients with T2D.	• Changes from baseline in patient-reported health status as measured by the EuroQol 5-dimension 5-level questionnaire (EQ-5D-5L) at 12 and 28. • Changes from baseline in patient-reported weight-related self-perception as measured by the Impact of Weight on Self-Perception Scale (IW-SP) at 12 and 28.

	• Assessments of patient perceptions regarding use and specific features of the delivery devices using the following modules of the Medication Delivery Device Assessment Battery (MDDAB): - Module 1: Single-Use Pen (SUP) Ease of Use Module (Week 6 and 28). - Module 2: SUP Device Features Module (Week 6 and 28). - Module 3a: SUP Experience Module (Week 6). - Module 3b: SUP Experience Module (Week 28). - Module 4: Glargine Delivery Device Experience Module (Week 28).

Abbreviations: CV = Cardiovascular; EQ-5D-5L = EuroQol 5-dimension 5-level questionnaire; FSG = fasting serum glucose; HbA1c = hemoglobin A1c; IW-SP = Impact of Weight on Self-Perception Scale; MDDAB = Medication Delivery Device Assessment Battery; BG = blood glucose; SMBG = self-monitored blood glucose; SUP = Single-Use Pen; T2D = type 2 diabetes; TTT = treat-to-target.

5. A Priori Statistical Methods

5.1. Determination of Sample Size

A total of approximately 290 patients (145 per group) will be randomized to have approximately 246 completers (123 per group) to show dulaglutide add-on titrated insulin glargine is superior to placebo add-on titrated insulin glargine with 90% power assuming a treatment difference of 0.5% in HbA1c reduction, standard deviation (SD)=1.2%, and dropout rate of 15%, and 2-sided alpha of 0.05. The screen failure rate is estimated as 40%. Approximately 483 patients will be screened.

5.2. General Considerations

This plan describes all priori statistical analyses for efficacy and safety that will be performed.

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

All analyses will be implemented using Version 9.4 SAS or higher.

Unless otherwise specified, listings will include all randomized patients. Both efficacy and safety analyses will be conducted using the intent-to-treat (ITT) population. Unless otherwise specified, efficacy data from patients who started rescue therapy or premature treatment discontinuation (whichever occurs first) will be censored from the point of initiating rescue treatments or early discontinuation onwards. Safety analysis will not censor this part of data. Selected analyses will also be conducted using the per-protocol (PP) and Completers populations. The PP and Completers populations are subsets of the ITT population.

Unless otherwise noted, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, and confidence intervals (CIs) will be calculated at 95%, 2-sided. All tests of interactions between treatment groups and other factors will be conducted at a 2-sided alpha level of 0.10.

The baseline data will be collected at Visit 3. For all variables, including HbA1c, if baseline data are missing, the last non-missing measurement taken prior to this visit will be used for the baseline measurement.

Patients will be randomized in a 1:1 ratio to (1) dulaglutide 1.5 mg QW add-on to insulin glargine or (2) placebo QW add-on to insulin glargine, with metformin and/or acarbose. Patients will also be stratified by HbA1c at Visit 1 (screening) (<8.5% vs. ≥8.5%) and OAM usage (metformin alone, acarbose alone, metformin and acarbose) at Visit 3 (randomization) to achieve

between-group comparability and to mitigate against confounding the effects of the treatments with severity of disease or treatment with metformin and/or acarbose.

The primary analysis model will be mixed-model for repeated measures (MMRM) analysis using restricted maximum likelihood (REML) with treatment, OAM usage (metformin alone, acarbose alone, metformin and acarbose), visit, and treatment-by-visit interaction as fixed effects, baseline HbA1c as a covariate, and patient as a random effect. An unstructured covariance structure will be used to model the within-patient variability and implicitly adjust for missing data. There will be no imputation of missing data for patients who have no postbaseline data.

If this analysis fails to converge, the following covariance structures will be tested in order:

- Toeplitz with heterogeneity
- Autoregressive with heterogeneity, by visit
- Compound symmetry with heterogeneous variances, by visit
- Toeplitz
- Autoregressive
- Compound symmetry without heterogeneous variances, by visit

The first covariance structure that converges will be used. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. Significance tests will be based on least-squares (LS) means using Type III hypotheses.

The percentage of patients achieving the target HbA1c of <7.0% will be analyzed using a longitudinal logistic regression with repeated measurements.

For continuous measures, summary statistics will include sample size (n), mean, SD, median, minimum (min), and maximum (max) for both the actual and the change from baseline measurements. Least-squares mean (LS Mean) and standard errors (SEs) derived from the model will also be displayed for the change from baseline measurements. Treatment comparisons will be displayed showing the treatment difference LS Mean and the 95% CIs for the treatment differences along with the p-values for the treatment comparisons.

For categorical measures, summary statistics will include sample size, frequency, and percentages. Fisher's exact test will be used.

5.3. Graphical Testing Scheme

To control type I error, a graphical testing scheme (Bretz et al. 2011) presented in **Error! Reference source not found.** will be used to compare treatments regarding selected secondary objectives once the primary objective has been achieved. The numbers in **Error! Reference source not found.** represent the fraction of alpha to allocate to the next hypothesis if the current null hypothesis is rejected.

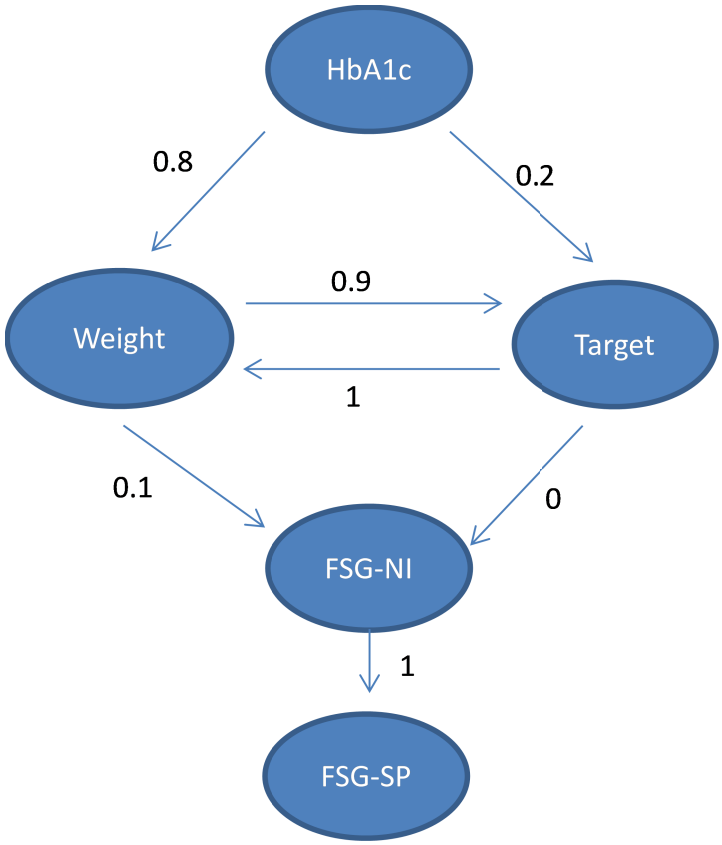

Figure 5.1 Graphical testing scheme for H9X-MC-GBGO.

Note:

HbA1c: primary objective, superiority test of the addition of dulaglutide 1.5 mg QW vs the addition of once weekly placebo to insulin glargine on HbA1c change from baseline at Week 28.

Weight: superiority test of the addition of dulaglutide 1.5 mg QW vs the addition of once weekly placebo to insulin glargine on body weight change from baseline at Week 28.

Target: superiority test of the addition of dulaglutide 1.5 mg QW vs the addition of once weekly placebo to insulin glargine on HbA1c target 7.0% at Week 28.

FSG-NI: noninferiority test of the addition of dulaglutide 1.5 mg QW vs the addition of once weekly placebo to insulin glargine on FSG change from baseline at Week 28. Noninferiority margin of 12mg/dL (equivalent to 0.4% of HbA1c margin) will be used.

FSG-SP: superiority test of the addition of dulaglutide 1.5 mg QW vs the addition of once weekly placebo to insulin glargine on FSG change from baseline at Week 28.

5.4. Patient Population

The following patient populations described in Table 5.1 will be used to analyze the data. The data collected in this study will be presented as listings including investigator site, patient, and treatment.

Table 5.1 Analysis Populations for Study H9X-MC-GBGO

Population	Definition
Entered	All patients who sign informed consent forms
Randomized	All patients who are randomized to a treatment arm
Nonrandomized	All patients entered, but not randomized to a treatment arm
Intent-to-Treat (ITT)	All patients randomized who take ≥1 dose of study medication for assigned treatment arm
Completers	All patients in the ITT population who complete the planned treatment period (that is, do not discontinue from the treatment early). Data after initiation of rescue therapy will be censored from the efficacy analyses. • Note: There is no requirement that completers have to be on study drug the entire study.
Per-Protocol (PP, see details in the document Trial Issue Management Plan)	All patients in the ITT population who also meet the following criteria: • Have no important protocol deviation expected to impact the primary endpoint for HbA1c (see below, Section 5.5.3); • Complete the planned treatment period (28 weeks [Visit 17]) for primary endpoint (that is, do not discontinue from the study drug early).

5.5. Treatment Group Comparability

5.5.1. Patient Disposition

A listing of patient discontinuation will be presented for all randomized patients. Frequency counts and percentages of all patients entered, randomized/enrolled, completing and discontinuing from the study and study treatment will be presented for both the treatment groups. A summary of discontinuations will also be presented by visit. The overall percent discontinued comparisons among the treatments will be performed using Fisher's exact test.

5.5.2. Patient Characteristics

Demographic and baseline clinical characteristics will be summarized by treatment group using ITT, PP and Completer populations. For continuous measures, summary statistics will include sample size (n), mean, SD, median, min, and max. Treatment differences will be analyzed using ANOVA model with treatment as the factor. For categorical measures, summary statistics will include sample size, frequency and percent. Treatment difference will be compared using Fisher's exact test.

5.5.3. Protocol Deviations

Important protocol deviations will be listed for all randomized patients. The rationale for choosing the important protocol deviations was based on their potential to impact the primary analysis.

A separate document (known as the Trial Issue Management Plan) will be used to define the categories, subcategories and study-specific term of important protocol deviations.

The listing of important protocol deviations (IPD) for all randomized patients during the entire study, with the indication of whether to be excluded from the PP population, will also be provided. The IPDs identified by site monitoring and clinical database will be integrated in the listing.

5.5.4. Concomitant Medications

Concomitant medications, including previous therapy for diabetes, will be summarized by treatment. Proportion of certain categories of concomitant medications will be analyzed using a Fisher's exact test.

5.5.5. Study Drug Compliance

Study drug compliance will be listed and summarized using ITT population. Study drug compliance is defined as taking at least 75% of the scheduled injections of the study drug between the visits of randomization (Visit 3, Week 0) and Week 4, 8, 12, 18, 24 and 28. The frequency and percent of patients who are compliant by treatment group will be summarized and compared using Fisher's exact test at these visits.

In addition, the overall compliance during the study will be calculated for each patient. The overall compliance in percentage for each patient will be calculated by taking the number of visits the patient was compliant divided by the total number of visits with non-missing compliance data for this patient *100. The overall compliance will be summarized and presented in descriptive statistics that include the sample size (n), mean, SD, median, min, and max. Overall treatment compliance for each patient is defined as taking at least 75% of the injectable study drug for at least 75% of the visits, that is, the overall compliance percentage is at least 75% for this patient.

5.5.6. Treatment Exposure

Treatment exposure is defined as the time from when a patient is randomized at Visit 3 and receives study drug until the patient either discontinues treatment or completes the treatment period.

The duration of treatment exposure will be summarized and listed by treatment group using ITT, PP, and Completer population. Duration of exposure will be categorized into the following groups: ≤14 days, >14 to ≤28 days, >28 to ≤56 days, >56 to ≤84 days, >84 to ≤126 days, >126 to ≤196 days and >196 days. These categories will be summarized as frequency counts by treatment group. Summary statistics will include mean per patient exposure in days, median, min, max and SD. The duration of treatment exposure will be analyzed using one-way ANOVA with treatment as fixed effects.

5.6. Primary Efficacy Analysis

The primary efficacy measurement in this study is change in HbA1c from baseline to Week 28, as determined by the central laboratory.

The primary analysis model will be MMRM for HbA1c change from baseline to Week 28 (Visit 17). The model will include treatment, OAM use, visit, treatment-by-visit interaction as fixed effects, and baseline HbA1c as a covariate and patient as a random effect. LS Mean, 95% CI and the p-value will be presented for the treatment difference (dulaglutide 1.5 mg minus placebo).

Primary analysis will be done using ITT population.

The MMRM analysis model for the primary endpoint will be repeated using the PP and Completer populations to check the sensitivity of the primary analysis.

5.7. Key Secondary Efficacy Analysis (Controlled for Type I Error)

5.7.1. Percentage of Patients Achieving a Target HbA1c <7.0%

Frequency counts and percentage of patients achieving the target of HbA1c <7.0% will be presented for the ITT population. The percentage of patients achieving the target of HbA1c <7.0% will be analyzed using longitudinal logistic regression with repeated measurements. The model will include independent variables of treatment, OAM usage, visit, and treatment-by-visit interaction as fixed effects and baseline HbA1c as a covariate.

5.7.2. Analysis of Body Weight and Fasting Serum Glucose (FSG)

Actual and change of weight and FSG from baseline will be summarized for the ITT population. The evaluation of the two dependent variables – change in body weight from baseline, and change in FSG from baseline, will be performed using MMRM model separately. The MMRM model will include treatment, OAM usage, visit, baseline HbA1c strata (<8.5%, and ≥8.5%), treatment-by-visit interaction as fixed effects, and baseline dependent variable as a covariate and patient as a random effect. LS Mean, 95% CI, and the p-value will be presented for the treatment comparison.

5.8. Additional Secondary Efficacy Analysis

5.8.1. Percentage of Patients Achieving a Target HbA1c ≤6.5% and the Composite Outcomes

The single endpoint (HbA1c of ≤6.5% at Week 28) will be analyzed using the same model described above for the analysis of patients achieving the target of 7%; double endpoint (HbA1c <7% at Week 28, and without documented symptomatic hypoglycemia [<3.0 mmol/L] during the maintenance period [Weeks 12-28]); double endpoint (HbA1c <7% at Week 28, and without

weight gain [<0.1 kg] at Week 28); and triple endpoint (HbA1c <7% at Week 28 and with no weight gain [<0.1 kg] at Week 28 and without documented symptomatic hypoglycemia [BG <3.0 mmol/L] during the maintenance period [Weeks 12-28],) will be analyzed by logistic regression. The model will include baseline HbA1c, OAM usage, and treatment as factors.

5.8.2. Analysis of 7-point SMBG profile

The 7-point SMBG profiles consist of pre-meal and 2-hour postprandial SMBG measurements for the morning, midday, and evening meals, and one measurement at bedtime. Patients will be asked to perform two 7-point SMBG profiles over a 24-hour period on 2 nonconsecutive days, in the 2-week period prior to prespecified office visits (Visits 3 [Week 0], 12 [Week 12], 17 [Week 28], and early termination [ET]). In general, the mean of the two daily values will be used for reporting purpose. If 7-point SMBG is not performed, then data from the most recent nonconsecutive 4-point SMBG (3 before each meal and one at bedtime) profiles can be used. If more than 2 SMBG profiles are available, the 2 most recent profiles on nonconsecutive days should be used.

The following variables for 7-point SMBG profile will be analyzed using MMRM model:

1. Pre morning meal BG (mg/dL)
2. 2-hour postprandial measurement for morning meal BG (mg/dL)
3. Pre midday meal BG (mg/dL)
4. 2-hour postprandial measurement for midday meal BG (mg/dL)
5. Pre evening meal BG (mg/dL)
6. 2-hour postprandial measurement for evening meals BG (mg/dL)
7. Bedtime BG (mg/dL)
8. Mean of all pre-meal BG (mg/dL)
9. Mean of all postprandial BG (mg/dL)
10. Mean of all meals 2-hour excursion (mg/dL)
11. Mean of all 7-point BG (mg/dL)

For before and 2 hours after meal BG and bedtime BG (Reports 1-7), the mean BG value at each visit is calculated as the average of 2 daily BG values. The change from baseline is calculated as the mean BG value at Week 12 and Week 28 minus the mean BG value at the baseline.

For mean of all pre-meal BG (Report 8), the pre-meal daily mean is calculated as the average BG values collected for before morning, midday and evening meals on a particular day. The mean of all pre-meal BG at each visit is calculated as the average of 2 pre-meal daily means. The change from baseline is calculated as the mean of all pre-meal BG at Week 12 and Week 28 minus the mean of all pre-meal BG at baseline.

For mean of all postprandial BG (Report 9), the post-meal daily mean is calculated as the average of 2-hour postprandial BG values of morning, midday and evening meals on a particular day. The mean of all postprandial meals BG at each visit is calculated as the average of 2 post-

meal daily means. The change from baseline is calculated as the mean of all postprandial BG at Week 12 and Week 28 minus the mean of all postprandial BG at baseline.

For mean of all meals 2-hour excursion at each visit (Report 10), the daily mean for all meals is calculated as the average of glucose excursion for morning, midday and evening meals on a particular day. The mean of all meals 2-hour excursion at each visit is calculated as the average of 2 daily means. The change from baseline is calculated as the mean of all meals 2-hour excursion at Week 12 and Week 28 minus the mean of all meals 2-hour excursion at baseline.

For mean of all 7-point BG (Report 11), the daily mean is calculated as the average of 7 BG values collected on a particular day. The mean of all 7-point BG at each visit is calculated as the average of 2 daily means. The change from baseline is calculated as the mean of all 7-point BG at Week 12 and Week 28 minus the mean of all 7-point BG at baseline.

Actual and change from baseline of 7-point SMBG will be summarized using ITT population, and analyzed using MMRM with treatment, OAM usage, visit, baseline HbA1c strata (<8.5%, and ≥8.5%) and treatment-by-visit interaction as fixed effects, baseline 7-point SMBG as covariate and patient as a random effect. An unstructured covariance structure will be used to model the within-patient variability and implicitly adjusts for missing data. If this analysis fails to converge, then other covariance structures will be tested in the order mentioned in Section 5.2.

5.8.3. Analysis of Insulin Glargine Dose

Actual and change of daily insulin glargine dose from baseline measurements will be summarized using ITT population. Change from baseline of glargine dose will be analyzed using MMRM model and will include treatment, OAM usage, visit, baseline HbA1c strata (<8.5%, and ≥8.5%), treatment-by-visit interaction as fixed effects, and baseline glargine dose as a covariate. Daily insulin dose of each visit is defined as the dose of the last dose before each visit.

5.9. Tertiary/Exploratory Efficacy Analyses

5.9.1. Insulin Glargine Treat-to-Target (TTT) Algorithm Assessment

The number of insulin dose assessments per the TTT algorithm, and compliance with the TTT algorithm at Week 6, 8 and 12 (number of assessments performed correctly; number of assessments that required dose change; number of assessments for which the outcomes were correctly followed by the patient) and frequency of insulin dose adjustments will be analyzed using an ANCOVA model including treatment, OAM usage, and baseline HbA1c strata. Reasons for patients not following the outcome of the assessment will be analyzed using Fisher's exact test.

5.9.2. Other Composite Outcomes

The other objectives include triple endpoint (HbA1c <7% and no weight gain [<0.1 kg] at Week 28, and no documented symptomatic hypoglycemia [BG < 3.9 mmol/L] during the maintenance

period); double endpoint (HbA1c <7% at Week 28 and no documented symptomatic hypoglycemia [BG < 3.9 mmol/L] during the maintenance period), and double endpoint (HbA1c <7% and with no weight gain [<0.1 kg] at Week 28) will be analyzed by the same model as that in Section 5.8.15.7.1.

5.10. Efficacy Analysis without Censoring Post-Rescue or Early-Discontinuation Data

We will conduct some additional efficacy analysis without censoring the data after rescue therapy or premature treatment discontinuation (whichever occurs first). Missing endpoints will be imputed by jumping to the reference (J2R) method as described in Carpenter et al. 13, which assumed that the drug effect for missing data in both the placebo arm and dulaglutide arm was like the observed effect in the placebo arm. This approach in essence assumed that subjects would lose any treatment effect after the intercurrent events, which might be used as a worst-case scenario in terms of assuming patients with missing data received no drug benefit after the intercurrent events. Measurements include actual and change from baseline in HbA1c, body weight and FSG and frequency counts and percentage of patients achieving the target of HbA1c <7.0%. We will analyze HbA1c, body weight and FSG by ANCOVA model. For HbA1c, treatment and OAM usage are fixed effects and baseline is a covariate. For body weight and FSG, treatment, OAM usage, baseline HbA1c strata (<8.5%, and ≥8.5%) are fixed effects, and baseline dependent variable is a covariate. Percentages of patients achieving the target of HbA1c <7.0% will be analyzed using logistic regression with treatment and OAM usage as fixed effects and baseline HbA1c value as a covariate.

5.11. Safety Analyses

The safety analysis will include analysis of adverse events (AEs), hypoglycemic and hyperglycemic episodes, vital signs, and laboratory analytes. Unless otherwise specified, the ITT population will be used for analyses of safety measures.

5.11.1. Adverse Events

An AE is any untoward medical event associated with the use of a drug in humans, whether or not it is considered related to a drug. Adverse events will be coded from the actual term described by the investigator using the Medical Dictionary for Regulatory Activities (MedDRA). Unless otherwise specified, AEs will be reported using the MedDRA system organ class (SOC) and preferred term (PT). Selected AEs may be reported using MedDRA high-level terms (HLT).

The following SMQs will be used to identify CV events:

1. Ischaemic heart disease SMQ
    a. Myocardial infarction SMQ (20000047)
    b. Other ischaemic heart disease SMQ (20000168)
2. Central nervous system haemorrhages and cerebrovascular conditions SMQ

   a. Conditions associated with central nervous system haemorrhages and cerebrovascular accidents SMQ (20000166)
   b. Haemorrhagic central nervous system vascular conditions SMQ (20000064)
   c. Ischaemic central nervous system vascular conditions SMQ (20000063)
3. Cardiac failure SMQ (20000004)

The following Standardized MedDRA Queries (SMQs) and customized Lilly Search Categories (LSC) will be used to identify Acute Pancreatitis (SMQ) and Chronic Pancreatitis (LSC):

- Acute Pancreatitis SMQ
  - Apply the Acute Pancreatitis SMQ (20000022), narrow terms only
- Chronic Pancreatitis LSC
  - Search for following PTs
    1. Pancreatitis chronic (10033649)
    2. Pancreatic duct obstruction (10052765)
    3. Pancreatic failure (10079281)
    4. Pancreatic enlargement (10055024)
    5. Pancreatic duct dilatation (10064858)
    6. Sphincter of Oddi dysfunction (10066424)
    7. Ampulla of Vater stenosis (10067376)
    8. Pancreatic calcification (10068823)

The following LSC will be used to identify Thyroid Neoplasm:

1. Thyroid neoplasms (10043747)
2. Thyroid neoplasms malignant (10043749)
3. Thyroid disorders NEC (10043712)
   a. 2 PTs only: Thyroid C-cell hyperplasia (10070568) and hypercalcitoninaemia (10072001)

The following SMQs will be used to identify Allergic and Hypersensitivity Reactions:

- Apply the following SMQs, narrow terms only
    1. Hypersensitivity SMQ (20000214)
    2. Anaphylactic Reaction SMQ (20000021)
    3. Angioedema SMQ (20000024)

The following LSC will be used to identify Injection Site Reactions:

- Find PTs nested within the HLT of Injection Site Reactions (10022097). Separately look at the following terms (a subset from the list above) which are potentially immune-mediated injection site reactions:
  1. Injection site dermatitis
  2. Injection site erosion
  3. Injection site erythema

4. Injection site hypersensitivity
5. Injection site induration
6. Injection site pruritus
7. Injection site exfoliation
8. Injection site vasculitis
9. Injection site photosensitivity reaction
10. Injection site rash
11. Injection site urticaria
12. Injection site hypertrophy
13. Injection site inflammation
14. Injection site irritation
15. Injection site oedema
16. Injection site warmth

For hepatic safety, apply the Drug related hepatic disorders - comprehensive search SMQ (20000006).

All AEs and treatment-emergent adverse events (TEAEs), defined as an event that first occurs or worsens (increases in severity) on or after first dose will be listed by patient and visit using MedDRA PT and SOC. Information on treatment, preferred term, reported term, severity, seriousness, and relationship to study drug will be reported. Listings of patients with study and study drug discontinuation due to AEs or death will be produced.

Summary statistics will be provided for TEAEs, SAEs, and study and study drug discontinuation due to AEs or death during the treatment period and through safety follow-up. Counts and proportions of patients experiencing AEs will be reported for each treatment group, and Fisher's exact tests will be used to compare the treatment groups.

Listings and summary statistics will also be provided for adverse event of special interest (AESI) including deaths, hypoglycemia, severe, persistent hyperglycemia, nonfatal CV events, adjudicated and confirmed pancreatitis, thyroid neoplasm, allergic and hypersensitivity reactions.

Since gastrointestinal (GI) AEs, like diarrhea, nausea, abdominal distension and vomiting, are among the most common events reported in patients treated with dulaglutide, summaries and analyses for onset, duration, and severity of diarrhea, nausea, abdominal distension and vomiting will be provided. The duration of an adverse event was calculated as adverse event end date minus the adverse event start date plus 1 day and was reported in terms of median number of days. Time to onset of an adverse event was analyzed by day, from the first dose date of the study drug to the occurrence of the adverse event.

The following PTs will be used to identify GIAEs:

1. Specific diarrhea symptoms include the following PTs
    a. Diarrhea (10012735)
    b. Fecal volume increased (10049939)

   c. Frequent bowel movements (10017367)
2. Specific nausea symptoms include the following PTs
   a. Retching (10038776)
   b. Procedural nausea (10066962)
   c. Nausea (10028813)
3. Special abdominal distension symptoms include the following PTs
   a. All PTs nested within the HLT of flatulence, bloating and distension (10016770)
4. Specific vomiting symptoms include the following PTs
   a. Retching (10038776)
   b. Vomiting (10047700)
   c. Vomiting projectile (10047708)

5.11.2. Hypoglycemic Episodes

Level 1, 2, 3, and nocturnal hypoglycemia will be included for summary and analysis. Classification of these hypoglycemic events are shown in Table 5.2.

Table 5.2 Definitions of Hypoglycemic Event Categories

Glycemic criteria/description	
Level 1 (Glucose Alert Value)	Blood Glucose <70 mg/dL (3.9 mmol/L) and >= 54 mg/d (3.0 mmol/L)
Level 2 (Clinically Significant Hypoglycemia)	Glucose <54 mg/dL (3.0 mmol/L)
Level 3 (Severe Hypoglycemia)	A severe event characterized by altered mental and/or physical status requiring assistance for recovery (no specific glucose threshold).

Nocturnal hypoglycemia: a hypoglycemia event (including severe hypoglycemia) that occurs between bedtime and waking.

A listing of the individual hypoglycemic episodes will be presented using all randomized population. Summary reports will include both incidence and rates of hypoglycemia for each visit including safety follow-up, as well as in stabilization period (Week 0 – 4), maintenance period (Week 12 – 28) and overall period (Week 0 – 28) will be reported for the ITT population without post-rescue or early-discontinuation data.

The incidence of hypoglycemic episodes will be summarized by frequencies and proportion for each treatment group using ITT population, by visit, stabilization period, maintenance period as well as overall period. The overall frequency and proportion are calculated as the total number of patients and proportion of patients reporting hypoglycemic episodes through safety follow-up. The maintenance frequency and proportion are calculated as the total number of patients and

proportion of patients reporting hypoglycemic episodes during Weeks 12 to 28. Treatment differences in the incidence of hypoglycemic episodes will be assessed using Fisher's exact test.

The rates of hypoglycemic episodes (episodes/patient/year) for each visit, overall and maintenance period will be analyzed using a generalized linear mixed-effects model assuming negative binomial distribution. The model will include treatment, OAM usage (metformin alone, acarbose alone, metformin and acarbose), baseline HbA1c strata (<8.5% vs. ≥8.5%), as fixed effects. The logarithm of days between visits will be included as an offset to account for possible unequal duration between visits and between patients.

Three additional categories of hypoglycemia will be used in efficacy analysis in Section 5.8.1 and 5.9.2. The definitions are below:

Total hypoglycemia: all level 1/2/3 events combined.

Documented symptomatic hypoglycemia with BG <3.0 mmol/L (or 3.9 mmol/L): any time a patient felt that s/he was experiencing symptoms and/or signs associated with hypoglycemia and had a BG <3.0 mmol/L (or 3.9 mmol/L).

Asymptomatic Hypoglycemia: an event not accompanied by typical symptoms of hypoglycemia, but with a measured BG <3.9 mmol/L.

5.11.3. Severe, Persistent Hyperglycemia

Frequency counts and percentage of patients initiating rescue therapies due to severe persistent hyperglycemia will be presented using ITT population. Fisher's exact test will be applied to compare treatments on the proportion of patients initiating rescue therapies due to severe persistent hyperglycemia. Time to initiating rescue therapies due to severe persistent hyperglycemia will be analyzed between treatment groups using semi-parametric proportional hazard Cox model with treatment and OAM usage as fixed effects, and baseline HbA1c as covariate. A Kaplan-Meier curve will be plotted for both treatments on the same graph, for presentation purposes. Listing of patients who experienced severe, persistent hyperglycemia will be provided.

5.11.4. Vital Signs

Vital signs (systolic blood pressure, diastolic blood pressure, and pulse rate) and their change from baseline will be summarized by descriptive statistics (mean, SD, median, standard error, min, and max) by treatment group and visit for the ITT population. The average values of the 2 measurements of each vital sign will be used for the analysis.

In addition, the MMRM model will be applied to evaluate the change from baseline in vital signs as the dependent variable with treatment, OAM usage, visit, baseline HbA1c strata (<8.5%, and ≥8.5%) and treatment-by-visit interaction as fixed effects, baseline vital signs as a covariate and patient as a random effect. The LS Means will be estimated using the observed margins (OM) option in SAS MIXED procedure.

The number and percentage of patients with treatment-emergent outlier vital signs at each post-baseline visit will be presented and analyzed using Fisher's exact test for systolic blood pressure, diastolic blood pressure, and pulse rate according to the categories as listed below:

- Low systolic blood pressure (≤90 mmHg and a decrease from baseline ≥20 mmHg)
- High systolic blood pressure (≥140 mmHg and an increase from baseline ≥20 mmHg)
- Low diastolic blood pressure (≤50 mmHg and a decrease from baseline ≥10 mmHg)
- High diastolic blood pressure (≥90 mmHg and an increase from baseline ≥10 mmHg)
- Low pulse rate (<50 bpm and a decrease from baseline ≥15 bpm)
- High pulse rate (>100 bpm and an increase from baseline ≥15 bpm)

5.11.5. Electrocardiogram

Count and percent of patients with a change in the ECG from baseline and with an ECG changes which is an adverse event will be analyzed using Fisher's exact test.

5.11.6. Analysis of Laboratory Analyte

A listing of all laboratory measurements including scheduled and unscheduled, by site and by visit, using all randomized patients, will be provided. A listing of all laboratory measurements that are outside the normal range will also be presented for. Certain laboratory measurements will be listed using clinical relevant thresholds other than laboratory limits.

All summary analyses will be conducted by treatment group, using ITT population. Descriptive statistics will be presented, by treatment group and visit, for the laboratory measurements.

For each continuous laboratory measurement, the change from baseline, at endpoint will be analyzed using ANOVA on ranks, with treatment as a fixed effect. Last-observation-carried-forward (LOCF) values will be used to impute missing post baseline values for the last visit. The incidence and percent of high, low and normal values will be listed for each of the treatment arms and compared using the Fisher's exact test. Shift tables for selected analytes using clinically meaningful thresholds will be summarized descriptively.

For subjective (qualitative) laboratory assessments, count and percent of normal and abnormal values will be analyzed using Fisher's exact test.

Counts and percentages for patients with pancreatic enzymes above upper limit of normal (ULN) and greater than or equal to 3 x ULN will be summarized.

A summary of changes (shift tables using normal and abnormal categories) in amylase (total) and lipase evaluation from baseline to the maximum postbaseline will be produced.

Additional analyses will be conducted if deemed necessary.

5.11.7. Pharmacokinetic/Pharmacodynamic Analyses

Not applicable for this study.

5.12. Subgroup Analyses

Population subgroups of interest will be analyzed for the variable of HbA1c.

The following are candidate subgroups that might be analyzed. This list is not necessarily all-inclusive:

- Sex
- Baseline age group (<65 years, ≥65 years)
- Duration of diabetes at baseline (<median duration and ≥median duration)
- Baseline body mass index (<median and ≥median)
- Baseline HbA1c strata (<8.5%, ≥8.5%)
- Baseline FSG (by quartile)
- Baseline insulin dose (<median and ≥median)
- OAM usage (metformin alone, acarbose alone, metformin and acarbose), if more than 20 patients in each subgroup stratum.

An analysis will be performed examining the interaction term of the treatment by subgroup at Week 28 using the MMRM model with treatment, OAM usage, visit, subgroup, subgroup-by-treatment interaction, treatment-by-visit interaction, subgroup-by-visit interaction, and subgroup-by-treatment-by-visit interaction as fixed effects, baseline HbA1c as a covariate and patient as a random effect. For the subgroup analysis of baseline HbA1c strata (<8.5%, ≥8.5%), the model will include treatment, OAM usage, visit, subgroup, subgroup by treatment interaction, treatment-by-visit interaction, subgroup-by-visit interaction, and subgroup-by-treatment-by-visit interaction as fixed effects. Baseline HbA1c will not be included as a covariate. The interaction effect will be evaluated using a significance level of 0.10, unadjusted. If the subgroup is one of the stratification variables, then the subgroup will only be included once in the model.

5.13. Patient-Reported Outcome Analyses

The exploratory patient-reported outcome (PRO) analyses will be analyzed using the ANCOVA model with fixed effects treatment, OAM usage, baseline HbA1c strata (<8.5%, ≥8.5%) and baseline score as a covariate, using on-treatment without rescue data. LOCF will be used to impute missing postbaseline values, and treatment contrasts will be computed.

Questionnaire-specific analyses are summarized as follows:

- EQ-5D-5L: Summary statistics, including number of patients and proportion of categorical outcomes (5 levels) for the 5 dimensions (mobility, self-care, usual activity, pain/discomfort, and anxiety/depression) of EuroQol 5-dimension 5-level (EQ-5D-5L) questionnaire will be provided by treatment, at baseline, Week 12 and Week 28. Summary statistics of actual value and the changes from baseline to Week 12 LOCF endpoint and Week 28 LOCF endpoint in the EQ-5D-5L UK and Chinese population-based health state index score and EuroQol visual analog scale (EQ-VAS) score will be

summarized. EQ-5D-5L Chinese Population-based index score (Luo et al. 2017) can be derived to produce a patient-level index score between 0 and 1: https://ars.els-cdn.com/content/image/1-s2.0-S1098301516341250-mmc1.xls. The changes from baseline will be analyzed using the ANCOVA model for UK and China Index score and VAS score.

- IW-SP: Total score will be calculated, and mean, SD, min, and max scores for each domain will be provided by visit and by treatment. Mean, SD, median, min, and max scores will be provided for changes from baseline to Week 12 (Visit 12) and Week 28 (Visit 17). An ANCOVA model for the changes from baseline to Weeks 12 and 28 will be conducted. Significance of changes within each treatment arm and significance of difference between treatment groups will be reported.

The modules within the MDDAB (Modules 1, 2, 3a, 3b, and 4) will be evaluated as exploratory objectives in this study. The planned reports for each module by visit are as follows:

- Module 1, SUP ease of use: Each item in the module will be summarized using descriptive statistics (number of patients, mean, SD, median, first quartile, third quartile, min, and max) and frequencies and percentages of responses for each of the 5 categorical response options. Frequencies and percentages of patients who responded “Easy” or “Very Easy” will be summarized for each item.
- Module 2, SUP device features: Each item in the module will be summarized individually using descriptive statistics (number of patients, mean, SD, median, first quartile, third quartile, min, and max) and frequency counts and percentages of responses for each of the 5 response options. Additionally, frequency counts and percentages of patients who responded “Like” or “Extremely Like” will be summarized for each item.
- Modules 3a and 3b, SUP experience modules (3a: Visit 8 [Week 6] assessment; 3b: Week 28 and ET assessment): each item will be summarized individually using descriptive statistics (number of patients, mean, SD, median, first quartile, third quartile, min, and max) and frequency counts and percentages of responses for each of the response options. Additionally, frequency counts and percentages of patients who responded “Strongly Agree” or “Agree” (Items 1-3); “Confident,” “Very Confident,” or “Extremely Confident” (Items 4-6); “Mostly Willing” or “Definitely Willing” (Items 7-8) will be summarized.
- Module 4, Glargine delivery device experience module, will be summarized in the same way as module 3b (SUP experience module for end of study and ET). No formal statistical testing will be conducted to compare experience with the SUP at end of study (Module 3b) and experience with the glargine delivery device at end of study (Module 4); the study is not designed for comparative device evaluation, as patients will have uneven exposure to devices, both in terms of duration of use and frequency of administration.

6. Unblinding Plan

6.1. Interim Analyses

No interim analyses are planned for this study. If an unplanned interim analysis is deemed necessary, the appropriate Lilly medical director will be consulted to determine whether it is necessary to amend the protocol.

6.2. Site Level Unblinding

To preserve the blinding of the study, a minimum number of Lilly personnel will see the randomization table and treatment assignments before the study is complete. The treatment assignments will be blinded to patients and investigators until the end of the study.

Emergency un-blinding for AEs may be performed through the interactive web response system (IWRS). This option may be used ONLY if the patient's well-being requires knowledge of the patient's treatment assignment. All calls/ website visits resulting in an un-blinding event are recorded and reported by the IWRS.

The investigator should make every effort to contact the Lilly clinical research physician (CRP) prior to un-blinding a patient's treatment assignment. If a patient's treatment assignment is un-blinded, Lilly must be notified immediately by telephone.

If an investigator, site personnel performing assessments, or patient is un-blinded, the patient must be discontinued from the study. In cases where there are ethical reasons to have the patient remain in the study, the investigator must obtain specific approval from a Lilly CRP for the patient to continue in the study.

6.3. Sponsor/Trial Level Unblinding

The study team will remain blinded to treatment assignments until all patients have completed the study and the database has been finalized and locked for analysis.

7. References

American Diabetes Association. Standards of medical care in diabetes—2012. *Diabetes Care.* 2012;35:S11-S63.

Bretz F, Maurer W, Hommel G. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. *Statistics in Medicine*. 2011; 13:1489–1501.

International Diabetes Federation Clinical Guidelines Task Force. Global guideline for Type 2 diabetes: recommendations for standard, comprehensive, and minimal care. *Diabet Med*. 2006;23(6):579-593.

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

Luo, N., Liu, G., Li, M., Guan, H., Jin, X., & Rand-Hendriksen, K. (2017). Estimating an EQ-5D-5L value set for China. *Value in Health*, 20(4), 662-669.

Signature Page for VV-CLIN-066997 v1.0

Approval	PPD 22 02:07:47 GMT+0000

Signature Page for VV-CLIN-066997 v1.0