H9X-MC-GBGC Statistical Analysis Plan Version 7

A Randomized, Double-Blind Study Comparing the Effect of Once-Weekly Dulaglutide with Placebo in Pediatric Patients with Type 2 Diabetes Mellitus (AWARD-PEDS: Assessment of Weekly AdministRation of LY2189265 in Diabetes-PEDiatric Study)

NCT02963766

Approval Date: 18-Feb-2021

1. Statistical Analysis Plan: H9X-MC-GBGC: A Randomized, Double-Blind Study Comparing the Effect of Once-Weekly Dulaglutide with Placebo in Pediatric Patients with Type 2 Diabetes Mellitus (AWARD-PEDS: Assessment of Weekly AdministRation of LY2189265 in Diabetes-PEDiatric Study)

dulaglutide (LY2189265) T2DM

Study H9X-MC-GBGC is a Phase 3, randomized, double-blind, placebo-controlled trial with an open label extension that investigates the effect of the addition of dulaglutide (0.75 and 1.5 mg/week) or placebo weekly to metformin and/or basal insulin on change from baseline in hemoglobin A1c at 26 weeks in children and adolescents with type 2 diabetes mellitus.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol H9X-MC-GBGC
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 22-Jun-2015.
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 25-Jun-2015.
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly: 25-Aug-2015.
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly: 27-Jul-2016.
Statistical Analysis Plan Version 5 electronically signed and approved by Lilly: 04-Nov-2016.
Statistical Analysis Plan Version 6 electronically signed and approved by Lilly: 24-Jan-2018.
Statistical Analysis Plan Version 7 electronically signed and approved by Lilly on date provided below.

Approval Date: 18-Feb-2021 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

List of Figures

Figure **Page**

3. Revision History

Version 1 of this statistical analysis plan (SAP) was approved before the protocol was finalized for the purpose of sharing with the Food and Drug Administration (FDA) in a Type C meeting.

Version 2 reflected changes made to align the SAP with changes made to the draft protocol prior to sharing it with the FDA.

Version 3 of the SAP reflected changes made to address the FDA's feedback received from the above Type C meeting. More details were included regarding the sample size re-estimation (SSR) and analyses to address missing data.

Version 4 of the SAP reflected changes made to address the FDA and the European Medicines Agency's (EMA's) feedback regarding the study design and the statistical methods and aligned with the approved amended protocol H9X-MC-GBGC(a). It provided additional clarity and some necessary modifications and addressed graphical approach testing scheme to control Type 1 error among multiple comparisons of interest. It was approved prior to the study initiation to maintain *a priori* status for the SAP.

The overall major changes and rationale for the changes incorporated in Version 4 are summarized below:

1. To reflect the changes made in the protocol:
    a. Added study design diagram, dulaglutide 0.75 mg and 1.5 mg pooled analysis in primary and key secondary objectives to evaluate change from baseline in hemoglobin A1c (HbA1c), body mass index (BMI), fasting blood glucose (FBG) to Week 26 and HbA1c target of ≤6.5% at Week 26 in Section 4, Sections 5.1 to 5.3, and Sections 5.11 and 5.12.
    b. Changed randomized subjects per arm from 75 to 50, replaced 4* with 6× in SSR formula, and added Bayesian approach as sensitivity analysis for SSR assessment in Section 5.1.
    c. Removed "patient as a random effect" from all mixed-effects model for repeated measures (MMRM) models and added appropriate estimate statement to form dulaglutide pooled arm effect and comparison between dulaglutide pooled and placebo arms; moved autoregressive as an option after Toeplitz to a new row in Section 5.2.
    d. Replaced gatekeeping testing strategy with graphical approach in Section 5.3.
    e. Added health outcome addendum population, and revised per protocol, completer set populations definition restricted to Week 26 in Section 5.4.
    f. Changed external statistical analysis center (SAC) to internal SAC but external to study team in Section 6.1.
2. Changed "intent-to-treat (ITT) estimand" to "treatment regimen estimand" in this entire document.
3. Added statistical analysis visit window rational and definition.
4. Rearranged Sections 5.11 and 5.12.

5. Added graphical approach testing scheme allowing further change after blinded evaluating the correlation matrix at Week 26 among change from baseline in HbA1c, FBG, and BMI in pediatric patient population prior to datalock in Section 5.3.
6. Added analysis for adverse events (AEs) including but not limited to AE, treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), hyperglycemia requiring rescue, vital signs, electrocardiogram (ECG), injection site reaction, allergic/hypersensitivity, and pubertal progression as a new Section 5.15. Added stratified Wilcoxon rank sum test as sensitivity analysis for hypoglycemia rate, which will be used as the primary analysis only if the negative binomial regression analysis fails to converge in Section 5.15.3. Re-organized analysis of laboratory analytes in details in Section 5.15.8.
7. Added all exploratory analysis in a stand-alone Section 5.16.
8. Added patient-reported outcome (PRO) analysis in Section 5.18.

Version 5 of the SAP reflects changes made to address the regulatory agencies' feedback on the approved protocol H9X-MC-GBGC(a) and SAP Version 4.0. The key changes are summarized below:

- EMA
  1. The primary analysis population for the primary objective for EMA is changed to the ITT population excluding those patients treated with diet and exercise only who are metformin naïve (Sections 5.3 and 5.11.1).
  2. Add ITT population by excluding metformin naïve patients for EMA in Section 5.4.
- FDA
  1. Delete paragraph regarding estimands and graphical approach from Section 5.2.
  2. Add Section 5.11.3 "Multiple Imputation for Missing Data" by moving "tipping point" approach to this section. Update "copy reference" using all nonmissing rescued subject data as reference to do imputation for all missing data within the same treatment. The observed nonmissing data and imputed data for the observed missing ones will be evaluated using analysis of covariance (ANCOVA) model. This result will be implemented in graphical approach for multiplicity adjustment (Sections 5.3).
  3. Add descriptive summary for the primary and key secondary endpoints using 14 years old as cutoff points (≤14, >14) and update age cutoff of "=" sign going with "<" in Sections 5.17 and 5.18.
  4. Adverse event of interest is reorganized by moving "hypoglycemia" to Section 5.15.2.1; add "Renal Impairment" summary under Section 5.15.2.6; and moving out "Anti-drug Antibodies (immunogenicity)," "major cardiovascular events," "malignant events," to Sections 5.15.3, 5.15.4, and 5.15.5, respectively.
  5. Add analyses for the pharmacodynamics biomarkers to evaluate insulin sensitivity and beta-cell function in Section 5.16.6. The pharmacodynamics secondary biomarker visit window is added as an entry into Table GBGC.8.1.

Version 6 of the SAP reflects changes made to update the graphical testing scheme, to add a criterion to exclude patients from the per protocol (PP) population, to remove the appendix Table GBGC.8.1 with the analysis windows, and to remove the Bayesian approach for sample size re-estimation. The key changes are summarized below:

1. Add the pooled arm comparison with placebo for FBG, BMI, and HbA1c target of ≤6.5% analyses in the graphical testing scheme since these are key secondary objectives. Rearrange the testing order, putting the testing of FBG before BMI, and BMI before HbA1c target of ≤6.5% to maximize the overall trial success rate.
2. Add 1 criterion to exclude patients from PP population: "Patients that use systemic glucocorticoids for more than 14 days," since this is a standard criterion to exclude patients from the PP population in adult dulaglutide studies.
3. Remove the appendix of the analysis windows. Will use the actual visit as collected.
4. Remove the Bayesian approach for the sample size re-estimation.

Version 7 of the SAP reflects changes made to add a primary database lock. The following key changes were made to align the SAP with amendments to the protocol or protocol addenda.

1. Updated key secondary, other secondary, and exploratory objectives.
2. Updated plasma glucose (PG) level for hypoglycemia from ≤70 mg/dL to <70 mg/dL for both safety analyses and exploratory efficacy analyses of composite endpoints.
3. Added that "Based on regulatory feedback, the sample size re-estimation was performed in 2019 and confirmed the original sample size estimation, suggesting that approximately 150 patients be enrolled."
4. Removed the completer set (CS) population and removed analyses for CS and PP populations.
5. Defined HbA1c strata in the models using baseline HbA1c.
6. Removed the following sensitivity analyses: ANCOVA with last observation carried forward (LOCF) and tipping point approach.
7. Removed the Chi-square test to use Fisher's exact test only.
8. For the treatment regimen estimand, changed the primary analysis model from MMRM to ANCOVA with multiple imputation (MI) for change from baseline in HbA1c, FBG and BMI; removed analyses without MI.
9. Removed the baseline HbA1c-by-visit interaction from the longitudinal logistic regression model for HbA1c target for efficacy estimand.
10. Changed the primary analysis model from longitudinal logistic regression to logistic regression for HbA1c target <7.0% for the treatment regimen estimand.
11. Replaced country subgroups with region subgroups (US vs non-US).
12. Added a primary database lock at 26 weeks; clarified that PK team external to the study team may gain access to the unblinded data during the double-blind phase prior to the primary database lock.

In addition, the following key changes were made to the SAP only:

13. Changed the primary analysis model from ANCOVA to MMRM for change from baseline in FBG for the efficacy estimand since FBG is a longitudinal measurement.
14. Updated the summary of related AEs to be for AEs related to the study treatment since AEs related to the study treatment instead of other reasons are more relevant to safety concerns.
15. Removed sections for clinical trial registry analyses since they are not relevant for the clinical study report (CSR).
16. Updated that the listing of patients with severe, persistent hyperglycemia will be generated for patients who received rescue therapy since investigators were to prescribe rescue therapy for this reason which is the data collected on the case report form.
17. Removed laboratory analyses by visit since this level of detail is not expected to be insightful.
18. Removed hypoglycemia analyses for asymptomatic, nocturnal, and probable hypoglycemia events since they are already included in total hypoglycemia analyses, and added an additional hypoglycemia analysis, which will report all hypoglycemia events with PG <54 mg/dL, regardless of symptoms, and all severe hypoglycemia episodes for the ITT population, including events that occurred after patients started rescue therapy, since this was requested by FDA in other studies.
19. Clarified that subgroup analyses will be performed for the treatment regimen estimand only.
20. For baseline HbA1c (≤8.0%, >8.0%) subgroup analysis, added back the baseline HbA1c to the model as a covariate so that the results for actual and change from baseline are consistent, and the model terms are consistent within subgroup analyses.
21. Streamlined PRO section to reflect the latest user guide.
22. Removed the section specifying major cardiovascular events as an AE of interest since it was not included in the protocol as such, and the reporting of these events in this study population is unlikely. Any cardiovascular events that do occur should be reported as AEs and will be included in the analysis of TEAEs.
23. Updated the graphical testing scheme based on relevance of endpoints to clinicians and simulations using blinded data and published results from other trials.
24. Based on FDA feedback, the copy reference approach for multiple imputation for the primary analysis based on the treatment regimen estimand was changed to the washout method.
25. Based on FDA feedback, an interaction term of treatment by baseline insulin use was added to the negative binomial model for the analysis of hypoglycemia rates. In addition, the analyses for the incidence of hypoglycemia episodes and hypoglycemia rates will be conducted separately for the subsets of patients using insulin versus those not using insulin at baseline.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to test the hypothesis that dulaglutide (0.75 mg and 1.5 mg, pooled) given subcutaneously once a week for 26 weeks to children and adolescents with type 2 diabetes mellitus (T2DM) who have inadequate glycemic control, despite diet and exercise, with or without metformin and/or basal insulin is superior to placebo in the treatment of T2DM, as measured by baseline to Week 26 change in HbA1c.

4.2. Secondary Objectives

The secondary objectives of the study are to assess the efficacy, safety, pharmacokinetics (PK) and pharmacodynamics (PD) of dulaglutide in patients.

The key secondary efficacy objectives are to compare dulaglutide 0.75 mg and dulaglutide 1.5 mg arms (individually and pooled) to placebo with respect to the following parameters:

- Change in HbA1c between baseline and Week 26 (individual doses only)
- Change in FBG between baseline and Week 26
- Percentage of patients with HbA1c <7.0% at Week 26 (ADA 2020)
- Change in BMI between baseline and Week 26

The other secondary efficacy objectives are to assess the 2 dulaglutide treatment groups (individually and pooled) with respect to the following parameters:

- Change in HbA1c between baseline and Week 52
- Change in FBG between baseline and Week 52
- Percentage of patients with HbA1c <7.0% at Week 52
- Change in BMI between baseline and Week 52

The secondary safety objectives are to compare the dulaglutide 0.75 mg and dulaglutide 1.5 mg arms (individually and pooled) to placebo with respect to the following parameters at 26 weeks, and to assess the following parameters in dulaglutide 0.75 mg and dulaglutide 1.5 mg arms (individually and pooled) at Week 52:

- Glucose management-related safety assessed by the incidence of self-reported hypoglycemic events and the incidence of patients requiring rescue for severe, persistent hyperglycemia.
- Pancreatic safety assessed by the incidence of cases of pancreatitis confirmed by adjudication and the effect on pancreatic enzymes.
- Thyroid-related safety assessed by the incidence of cases of thyroid TEAEs and effect on serum calcitonin.

- Immune system-related safety, including the incidence of dulaglutide anti-drug antibodies (ADAs) and the incidence of allergic and hypersensitivity reactions, and injection site reactions.

The secondary PK/PD objective is as follows:

- Characterization of the PK of dulaglutide and the relationship between dulaglutide exposure and key safety and efficacy measures.

4.3. Exploratory Objectives

The exploratory objectives are to compare the dulaglutide 0.75 mg and dulaglutide 1.5 mg arms (individually and pooled) to placebo with respect to the following parameters at 26 weeks, and to assess the following parameters in the dulaglutide 0.75 mg and dulaglutide 1.5 mg arms (individually and pooled) at Week 52 (unless otherwise specified):

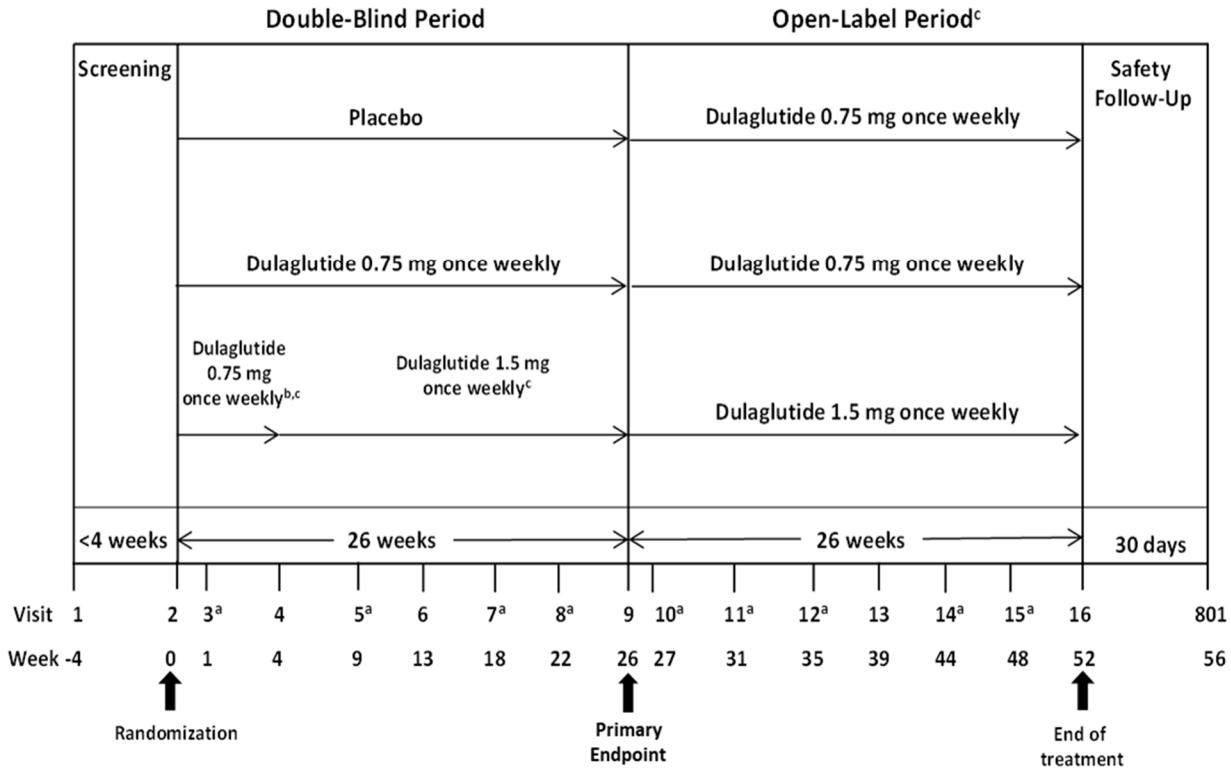
- Percentage of patients with HbA1c ≤6.5% at Weeks 26 and 52
- Change in HbA1c between baseline and Week 13
- Percentage of patients having HbA1c ≤6.5% without severe, documented symptomatic (<70 mg/dL), or probable hypoglycemic episodes at Weeks 26 and 52
- Percentage of patients having HbA1c <7.0% without severe, documented symptomatic (<70 mg/dL), or probable hypoglycemic episodes at Weeks 26 and 52
- Change in weight between baseline and Weeks 26 and 52
- Change in hormone-related safety assessed by the effect on morning serum prolactin, insulin-like growth factor-1 (IGF-1), estradiol, testosterone (males only), luteinizing hormone (LH), and cortisol between baseline and Weeks 26 and 52
- Change in serum lipids between baseline and Weeks 26 and 52
- Change in Tanner staging between baseline and Weeks 26 and 52
- Change in BMI standard deviation score (SDS) between baseline and Weeks 26 and 52
- Change in height and height SDS between baseline and Weeks 26 and 52
- Change in waist circumference between baseline and Weeks 26 and 52
- Change in the EQ-5D-Youth version (EQ-5D-Y) visual analogue scale (VAS) score between baseline and Weeks 26 and 52
- Percentage of patients reporting each level of problem on each dimension of the EQ-5D-Y at baseline and Weeks 26 and 52
- Assess effect of dulaglutide on measures of insulin resistance, beta cell function, and serum adiponectin at Weeks 13 and 26

- Change in basal insulin dose from baseline to Week 26 and from baseline to Week 52

5. A Priori Statistical Methods

Study H9X-MC-GBGC (GBGC) is a Phase 3, multicenter, randomized, double-blind, parallel-arm, placebo-controlled superiority trial with an open-label extension (Figure GBGC.5.1). A minimum of 150 male and female children and adolescents (ages 10 to <18 years) with T2DM and inadequate glycemic control on diet and exercise alone or diet and exercise plus metformin and/or basal insulin will be enrolled. Randomization will be stratified by the patient's background therapy and screening HbA1c. There will be a limitation on the number of patients with inadequate glycemic control managed by diet and exercise alone who have not previously received metformin so that these patients constitute no more than 25% of the total number of completers. The sponsor may place limitations on the patients enrolled depending on their demographics to meet regulatory expectations.

The main study has 4 periods: (1) a screening period lasting up to 4 weeks; (2) a double-blind treatment period lasting 26 weeks; (3) an open-label extension period lasting 26 weeks; and (4) a 30-day safety follow-up period.



a Phone visits.

b Patients who tolerate the 0.75-mg/week dose in the opinion of the investigator will be changed to dulaglutide 1.5 mg/week at Week 4.

c If patients do not tolerate the higher dose, they will be allowed to go back to the previous dose.

Figure GBGC.5.1. Illustration of study design for Study H9X-MC-GBGC.

5.1. Determination of Sample Size

Approximately 150 patients will be enrolled (50 patients per arm), in order to obtain 120 completers assuming a dropout rate of 20%. This provides at least 90% power for demonstrating superiority of the pooled dulaglutide 0.75 mg and 1.5 mg arms to placebo (primary objective) assuming a difference in mean change from baseline in HbA1c of -0.65% and standard deviation (SD) of 1%. Under the same assumptions, each individual dulaglutide arm will have at least 80% power to demonstrate superiority over placebo. The screen failure rate is estimated as 80%. Approximately 750 patients will be screened.

Because of the high degree of uncertainty regarding the appropriate sample size, the SAC may perform an SSR after approximately 100 patients have been enrolled. Based on treatment variability only (not treatment effect), the SAC may advise an increase in sample size up to approximately 189 patients enrolled (63 per arm) in order to obtain 150 completers. This new sample size would provide at least 80% power for demonstrating superiority of the pooled dulaglutide 0.75 mg and dulaglutide 1.5 mg arms to placebo with a difference in mean change from baseline in HbA1c of -0.65% and SD of 1.3% assuming a dropout rate of 20%. Under the same assumptions, each individual dulaglutide arm will have at least 69% power to demonstrate superiority to placebo.

Because of the anticipated difficulties in enrollment and the necessity for meeting regulatory timelines, Eli Lilly and Company (Lilly) will ultimately decide whether to conduct this SSR after taking into consideration the study enrollment rate and other factors. This calculation will not require input from the Data Monitoring Committee (DMC). The SAC will provide a brief communication to the Lilly Study Team indicating only the recommended sample size (up to approximately 189 patients). Lilly will then have the ability to adopt the recommended sample size or stick to the original sample size if it is determined to be too challenging to enroll or there are other feasibility concerns. No alpha adjustment is necessary for this analysis as it does not increase Type 1 error (Mehta and Tsiatis 2001; Pritchett et al. 2011).

To re-estimate the SD, a MMRM analysis of the change from baseline in HbA1c using restricted maximum likelihood (REML) will be used. This model will be fitted with baseline HbA1c as a covariate, stratification factors (insulin usage, metformin usage), treatment, visit, and treatment-by-visit interaction as fixed effects. An unstructured covariance structure will be used to model the variability. The Kenward-Roger method will be used to estimate the denominator degrees of freedom. The analysis model incorporates treatment and treatment-by-visit interaction terms to ensure accurate calculation of the total SD. However, the only output from this analysis will be the SD at Week 26 if variance-covariance is unstructured. No information on the observed treatment difference will be reported since it will not be utilized to determine the recommended sample size. The estimated SD is denoted as $\hat{\sigma}$. The dropout rate will not be reassessed during the assessment of the variability. The assumed value of 20% (0.2) in the initial power calculation will be used. The dropout rate is denoted as f. Based on the estimated variability and the assumed dropout rate, the recommended sample size, $\widehat{N}$, is estimated using,

$$\widehat{N} = 6 \times \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{(\mu/\widehat{\sigma})^2}\left(\frac{1}{1-f}\right)$$

with $\alpha = 0.025$ (one-sided), $\beta = 0.2$, $\mu = -0.65\%$, $f = 0.2$, and Z representing the standard normal quantiles.

The allowable range of the total sample size is constrained in a pre-specified manner, subject to an approximate minimum $N_{\min}$ of 150 and maximum sample size $N_{\max}$ of 189 patients. If the recommended sample size is greater than $N_{\max}$, then $N_{\max}$ will be used. If the estimated final sample size is less than $N_{\min}$, then $N_{\min}$ will be used. The SSR procedure was evaluated via simulation and is efficient in determining the appropriate sample size under a range of possible scenarios without inflating Type 1 error. No alpha adjustment to the final analysis is planned whether or not the sample size is increased.

Based on regulatory feedback, the sample size re-estimation was performed in 2019 and confirmed the original sample size estimation, suggesting that approximately 150 patients be enrolled. This figure is based on an estimated HbA1c SD of 1.4% provided by the SAC and assumed treatment difference of -0.8% and assumed proportion of patients with missing HbA1c measurement at 26 weeks equal to 10%.

5.2. General Considerations

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

All analyses will be conducted under the guidance and approval of Lilly statisticians.

The treatment groups mentioned in this document are dulaglutide (Dula) 1.5 mg and 0.75 mg (pooled); Dula 1.5 mg; Dula 0.75 mg; and Placebo during the double blinded period (randomization to Week 26) unless otherwise specified. Placebo becomes Placebo/Dula 0.75 mg during the open label period (Week 26 to Week 52) while the other treatment groups are unchanged.

For the purposes of tables, figures, and listings (TFLs), up to Week 26 the treatment groups will be All Dula (pooled dulaglutide doses); Dula 1.5; Dula 0.75; and Placebo. Placebo will be the reference treatment across all statistical reports whenever the treatment effect is compared between All Dula and Placebo, Dula 1.5 and Placebo, and Dula 0.75 and Placebo.

Up to Week 52 or to Week 56 (for selected safety parameters such as ADA at safety follow-up (Visit 801 [V801]), treatment will be expressed in the TFLs as All Dula, Dula 1.5, Dula 0.75, and Placebo/Dula 0.75 representing (Dula 0.75/0.75, Dula 1.5/1.5 pooled), Dula 1.5/1.5, Dula 0.75/0.75, and Placebo/Dula 0.75, respectively. No inference will be made among the treatment groups after Week 26 to Week 52 or from Week 26 to V801. Within-treatment comparisons

between different scheduled visits such as Week 26 and Week 52 for parameters such as HbA1c change (and others) from baseline may be performed.

Generally, V801 data will not be summarized unless otherwise stated, but V801 data may be listed.

Efficacy and safety data will be summarized by each treatment group at each scheduled visit unless otherwise stated.

Six patient populations are defined for the analyses in this study with detailed information listed in Table GBGC.5.1.

Unless otherwise specified, data listings will consist of all randomized patients. Both efficacy and safety analyses will be conducted using the ITT population. For safety except hypoglycemia analyses, all measurements including those made after taking rescue therapy will be included. Hypoglycemia analyses will be evaluated in the ITT population after censoring the data following administration of any rescue medication.

Unless otherwise specified, all tests of treatment effects will be conducted at a 2-sided alpha level of 0.05 and confidence intervals (CI) will be calculated at 95%, 2 sided.

For subgroup analyses, all tests of interactions between treatment and factors of interest will be conducted at a 2-sided alpha level of 0.10.

Visit 2 is the baseline visit. Baseline measurement for a corresponding parameter of interest is defined as the last nonmissing value taken prior to the first dosing of study drug. Baseline HbA1c value will be used to define HbA1c strata for analyses.

Unless otherwise specified, longitudinal continuous measures will be analyzed using MMRM with stratification factors (insulin usage [yes, no], metformin usage [yes, no], HbA1c strata [HbA1c: $<8.0\%$, $\geqslant 8.0\%$]), treatment (Dula 1.5, Dula 0.75, Placebo or Placebo/Dula 0.75), visit, and treatment-by-visit as fixed effects, and corresponding baseline measurement as a covariate unless otherwise noted. The Kenward-Roger method will be used to estimate denominator degrees of freedom, and the REML approach to obtain model estimate. An unstructured covariance structure will be used to model the within-patient errors. If this model 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 variance-covariance structure that converges will be used.

For continuous variables, summary statistics will include the number of patients, mean, SD, median, minimum, and maximum for both the actual and change from baseline for each treatment group at each visit. If a patient has no postbaseline data, the patient will be excluded from inferential statistical analysis. For data collected prior to first dosing, or randomization date if the first dosing date is not available, an analysis of variance (ANOVA) model with treatment (Dula 0.75, Dula 1.5, and Placebo) will be fitted. For post-first dosing longitudinal data, the aforementioned MMRM model will be fitted. For nonlongitudinal endpoint data, an ANCOVA model will be fitted by removing visit and visit-by-treatment interaction terms from its corresponding aforementioned MMRM model. Least-squares mean (LSmean) and standard errors derived from the corresponding statistical model will also be displayed for the change from baseline at each visit for each treatment group. Treatment comparisons at the same visit will be displayed showing the LSmean and the 95% CIs for the treatment differences (All Dula – Placebo; Dula 1.5 – Placebo; Dula 0.75 – Placebo) at the same visit along with the corresponding p-values up to Week 26. Appropriate estimate statements will be constructed in each aforesaid model (ANOVA, MMRM, or ANCOVA) to obtain a corresponding endpoint estimate for All Dula and the treatment comparison between All Dula and placebo at each visit of interest up to Week 26.

Unless otherwise stated, categorical variables will be summarized by sample size, frequency, and percentages for each treatment group at each visit. Fisher's exact test will be used to compare the difference among the 3 treatment groups and between All Dula and placebo.

For LOCF, the last nonmissing value will be imputed as the endpoint if the endpoint is missing. Any missing component of date and/or time may be imputed following Lilly standards.

Additional analyses may be performed even after datalock if deemed needed.

All statistical analyses will be conducted with R or SAS Version 9.4® or higher unless otherwise stated.

5.3. Graphical Approach to Adjust Multiplicity

To control overall Type 1 error, a graphical approach (Bretz et al. 2011) presented in Figure GBGC.5.2 will be used to compare the treatment effect among the pre-defined parameters of interest to address the selected key secondary objectives defined in the protocol at Week 26 for the ITT population only once the primary objective (H1) is achieved. In this figure, the numbers along the arrows represent the fraction of alpha from a null hypothesis, if it is rejected, to be passed to the next hypothesis. This graphical approach will be conducted separately for each of the following estimands, and each will be tested at the full significance level of 0.05:

- Treatment regimen estimand from the ITT population regardless of rescue status or treatment adherence coupled with the washout approach for multiple imputation.
- Efficacy estimand from the ITT population with censoring of the postrescue data.
- Efficacy estimand from the ITT population with censoring of the postrescue data and excluding lifestyle only/metformin naïve patients.

Figure GBGC.5.2 may be revised per the observed correlation matrix using blinded data among the parameters of interest at Week 26 prior to datalock.

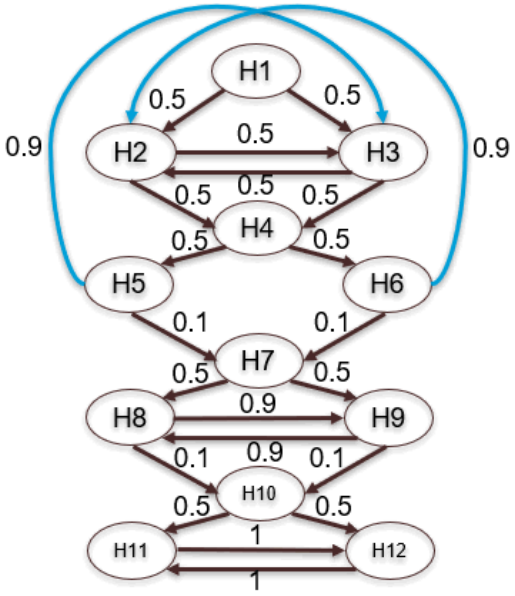


Figure GBGC.5.2. Graphical testing scheme for Study H9X-MC-GBGC.

- H_1: Superiority test of dulaglutide pooled arm (Pooled dulaglutide 1.5 mg and 0.75 mg) versus placebo in mean change from baseline in HbA1c at 26 weeks
- H_2: Superiority test of dulaglutide 1.5 mg versus placebo in mean change from baseline in HbA1c at 26 weeks
- H_3: Superiority test of dulaglutide 0.75 mg versus placebo in mean change from baseline in HbA1c at 26 weeks
- H_4: Superiority test of dulaglutide pooled arm versus placebo in proportion of patients achieving an HbA1c <7.0% at 26 weeks
- H_5: Superiority test of dulaglutide 1.5 mg versus placebo in proportion of patients achieving an HbA1c <7.0% at 26 weeks
- H_6: Superiority test of dulaglutide 0.75 mg versus placebo in proportion of patients achieving an HbA1c <7.0% at 26 weeks
- H_7: Superiority test of dulaglutide pooled arm versus placebo in mean change from baseline in FBG at 26 weeks
- H_8: Superiority test of dulaglutide 1.5 mg versus placebo in mean change from baseline in FBG at 26 weeks
- H_9: Superiority test of dulaglutide 0.75 mg versus placebo in mean change from baseline in FBG at 26 weeks
- H_{10}: Superiority test of dulaglutide pooled arm versus placebo in mean change from baseline in BMI at 26 weeks
- H_{11}: Superiority test of dulaglutide 1.5 mg versus placebo in mean change from baseline in BMI at 26 weeks

- H_{12}: Superiority test of dulaglutide 0.75 mg versus placebo in mean change from baseline in BMI at 26 weeks

5.4. Patient Population

The following patient populations described in Table GBGC.5.1 will be used to analyze the data.

Table GBGC.5.1. Analysis Populations for Study H9X-MC-GBGC

Population	*Definition*
All Entered	All patients who signed an informed consent
All Randomized	All patients who were randomized to a treatment arm as planned
Nonrandomized	All patients who entered, but were not randomized to a treatment arm
ITT	All randomized patients who took at least 1 dose of the study medication for an assigned treatment arm. For EMA, patients treated with diet and exercise alone who have not previously received metformin will be excluded from the ITT population for the primary analysis related to the primary endpoint (HbA1c) change from baseline to Week 26.
PP	All ITT population patients who have met the following criteria: • Have no important protocol deviations (Section 5.7) expected to influence the assessment of the primary objective • Have an overall compliance of at least 75% up to Week 26 • Have completed the double-blinded period (26 weeks therapy [Visit 9]) with nonmissing HbA1c for both the Week 26 and baseline visits
HO	A subset of the ITT population who signed an informed consent for the HO addendum and completed testing on at least 1 occasion.

Abbreviations: EMA = European Medicines Agency; HbA1c = hemoglobin A1c; HO = health outcomes; ITT = intent-to-treat; PP = per protocol.

5.5. Patient Disposition

Patient discontinuations will be listed and summarized for all randomized patients. The number and percentages of all patients who completed or discontinued from the study will be summarized for each treatment group from baseline to Week 26, from baseline to Week 52.

For the randomized population, discontinuations due to an AE or death, and discontinuations from study medication will be summarized separately for each treatment group.

The p-value to test the overall treatment effect and the p-value to compare the difference between All Dula and placebo across all reasons of discontinuation will be presented.

In addition, all entered patients will be summarized including, but not limited to, the total number of patients screened (entered), the number of patients excluded, the total number of patients randomized, the number of patients who entered into the double-blinded period, and the

number of patients who entered into the open-label period for each treatment group. Within each treatment group, the number of patients randomized but not treated, the number of patients included in the ITT population, the number of patients who completed 26 weeks of therapy, the number of patients who completed 52 weeks of therapy, the number of patients who completed safety follow-up (V801), and the number of patients discontinued will be summarized. Treatment difference will be compared using Fisher's exact test.

5.6. Patient Characteristics

Demographic and baseline clinical measures, including baseline Tanner stage by sex, will be summarized as described in Section 5.2 by each treatment group for the ITT and health outcome (HO) populations. For continuous measures, treatment differences will be analyzed using an ANOVA model with treatment as a fixed effect. For categorical measures, the overall treatment difference and between All Dula and placebo will be compared using Fisher's exact test. For either categorical or continuous measures, the p-value to test the overall treatment effect and the p-value to compare the difference between All Dula and placebo will be presented for each parameter of interest.

5.7. Important Protocol Deviations

Important protocol deviations will be listed for all randomized patients and summarized by treatment group. The rationale for choosing the important protocol deviations was based on their potential to impact the primary endpoint. The complete important protocol deviation list will be captured in an excel file, and then be converted in study data tabulation model (SDTM) as a source in addition to other necessary domains to define PP population. The decision and rationale for not reporting certain protocol deviations as important ones will be documented in the electronic trial master file (eTMF). The following protocol conditions will result in exclusion from the PP population:

- Informed consent was never obtained.
- Patients were randomized but the informed consent date is missing.
- Age <10 or age≥18 years old at randomization.
- Patients have known type 1 diabetes.
- Patients have a history of diabetic ketoacidosis after receiving antidiabetes medication.
- Patients have diabetes-associated autoantibodies (GAD65 or IA2), historically or at screening.
- Patients have a hemoglobinopathy or other disorder that interferes with the accurate determination of the primary endpoint, including, but not limited to, patients with the following hemoglobin variants: HbS, HbC, HbE, HbSC, and elevated HbF.

- Patients are taking a class of antihyperglycemic medication other than biguanides (metformin) and/or basal insulin during the first 26 weeks of the trial or have stopped these medications less than 3 months before Visit 1. These antihyperglycemic medications include but are not limited to a sulfonylurea, alpha-glucosidase inhibitor, DPP-IV inhibitor, GLP-1 RA, thiazolidinedione (TZD), glinide, and sodium-glucose co-transporter 2 (SGLT-2) inhibitor.
- Patients are taking prescription weight loss medication(s) chronically during the first 26 weeks of the trial or within 30 days of screening. These medications include, but are not limited to, the following: Contrave® (naltrexone/bupropion), Saxenda® (liraglutide), Xenical® (orlistat), Meridia® (sibutramine), Acutrim® (phenylpropanolamine), Sanorex® (mazindol), Adipex® (phentermine), BELVIQ® (lorcaserin), Qsymia™ (phentermine/topiramate combination), or similar over-the-counter medications (e.g., allī®).
- Patients have an HbA1c ≤6.5% or >11% if treated with antiglycemic medication; or >9% if treated with lifestyle only at Visit 1.
- Patients have a missing HbA1c at baseline (Visit 2) or at 26 weeks (Visit 9).
- Patients do not have an overall compliance with study drug of at least 75% up to 26 weeks.
- Patients that use systemic glucocorticoids for more than 14 days.

A complete list of important protocol deviations will be generated.

5.8. Concomitant Medications

Concomitant medications, including previous therapy for diabetes, will be summarized by different categories of medications and treatment groups using the ITT population. All concomitant therapies that were originally mapped using the World Health Organization (WHO) DRUG dictionary in the clinical trial database will be further classified using Anatomic-Therapeutic-Chemical (ATC) codes for reporting purposes.

Other prespecified concomitant medications of interest, including, but not limited to, antihypertensives, lipid lowering agents, and other cardiovascular agents; antidepressants and other psychiatric medications; antibiotics and other anti-infective therapies and other medications to treat the complications of diabetes will be summarized by treatment at baseline (Visit 2).

The number of patients and percentage of patients who took antihyperglycemic medication (nonrescued, and rescued) during the study will be summarized from Visit 1 to Visit 9 and from Visit 10 to Visit 16. Use of short-term (up to 2 weeks) of unlimited insulin therapy with continuation of study drug will not constitute rescue.

The p-value to test the overall treatment effect and the p-value to compare the difference between All Dula and placebo will be presented for each summary table.

The incidence of patients requiring rescue for severe, persistent hyperglycemia will be summarized in details in Section 5.15.5.

5.9. Study Drug Compliance

Study drug compliance will be listed and summarized for the ITT population. Patients with study drug change (dose change, study drug discontinuation) will be listed for all randomized patients.

A patient will be considered compliant for a period of interest if he or she is taking at least 75% of the expected doses for that period.

Overall treatment compliance from baseline to Week 26, and from baseline to Week 52 for each patient will be defined as taking at least 75% of the study drug for each period. The overall compliance in percentage for each patient will be calculated by taking the number of injections actually taken during that period divided by the total number of prescribed injections for this patient *100.

The p-value to test the overall treatment effect and the p-value to compare the difference between All Dula and placebo will be presented in a summary table.

The overall compliance defined as taking at least 75% prescribed medication from baseline to Week 26 will be used as one of the factors when determining if a patient is eligible for the PP population.

Overdosed subjects will be listed if those data are available.

5.10. Treatment Exposure

Treatment exposure is defined as the time from when the patient took his or her first dose until the last dosing date.

If that first dosing date is missing, it will be replaced with the randomized date. If the last dosing date is missing, it will be replaced with the date the patient discontinued treatment or completed the treatment period, whichever is earlier.

The duration of treatment exposure will be listed and summarized by each treatment group for the ITT population. The difference in exposure among treatment groups from baseline to Week 26, from Week 26 to Week 52, and from baseline to Week 52 will be compared using an ANOVA model with treatment as a fixed effect. The p-value to compare the overall treatment effect and the p-value to compare All Dula and placebo will be presented.

5.11. Primary Efficacy Endpoint Analyses

The primary efficacy measurement in this study is change in HbA1c from baseline to 26 weeks, as determined by the central laboratory.

5.11.1. Primary Analysis

Two primary estimands will be formed to compare the effect between All Dula and Placebo at Week 26. Each estimand will be tested with 2-sided alpha=0.05. The 2 estimands are derived from the ITT population for the primary efficacy endpoint:

1. Efficacy estimand from ITT population by censoring the post-rescue data
2. Treatment regimen estimand from ITT population regardless of rescue status or treatment adherence

The efficacy estimand measures the benefit of the assigned study treatment in the absence of the confounding effects of additional or alternative antihyperglycemic agents and regardless of compliance with study treatment. All scheduled post-first-dose-visit measurements up to Week 26 (Visit 9) inclusive will be used. The MMRM model mentioned in Section 5.2 will be fitted by replacing the baseline HbA1c strata with the baseline HbA1c as a covariate.

This method is chosen as the primary analysis as it is recommended by the National Research Council report (NRC 2010). This primary analysis is under missing at random (MAR) assumption and such an assumption often makes sense for the primary analysis (Little et al. 2012). Censoring the postrescue data minimizes the confounding rescue therapy's effect on the studied drug (O'Neill and Temple 2012).

The treatment regimen estimand measures the benefit of the assigned study treatment regardless of the use of any additional or alternative antihyperglycemic agents or compliance with study treatment. An ANCOVA model will be applied as the primary analysis to the complete data using multiple imputation (Section 5.11.3).

- Primary estimand for the FDA: treatment regimen estimand that will include post-rescue data
- Primary estimand for EMA: efficacy estimand excluding post-rescue data, and those patients treated with diet and exercise only who are metformin naïve from ITT population
- In general, for all other audiences: efficacy estimand that will not include post-rescue data (i.e., will censor the nonmissing, post-rescue data).

5.11.2. Sensitivity Analyses

The MMRM model mentioned in Section 5.11.1 may be applied using the efficacy estimand by replacing the fixed effect of the assigned treatment with the actual received treatment (i.e., those patients switched from the 1.5 mg dose to the 0.75 mg dose due to tolerability will be changed to the 0.75 mg group). The patients who switched from 1.5 mg to 0.75 mg will be listed.

5.11.3. Multiple Imputation for Missing Data

For the treatment regimen estimand, multiple imputation using the washout method which can be thought of as a modified placebo-based pattern mixture model (PMM) with a missing not at random (MNAR) assumption will be performed (NRC 2010; Permutt 2016). An ANCOVA with missing endpoints imputed using the washout method will be conducted. With the washout method, missing data from both treatment arms will be imputed using the data from those

patients in the placebo arm who had the measurement for the primary endpoint. Missing data from the dulaglutide arms will be imputed using only baseline and primary endpoint data from the placebo arm and none of the intermediate data observed in the placebo or dulaglutide arms. Missing data from the placebo arm will be imputed using both the baseline and all intermediate postbaseline and primary endpoint data in the placebo group. The aforementioned washout approach for missing data imputation will be implemented using SAS PROC MI and MIANALYZE. The ANCOVA model includes baseline HbA1c as a covariate, stratification factors, and treatment as fixed effects.

5.12. Key Secondary Efficacy Endpoint Analyses

The key secondary efficacy endpoints include change from baseline to Week 26 in HbA1c (each arm separately); the percentage of patients reaching a HbA1c target of <7% by Week 26; change from baseline to Week 26 for BMI and change from baseline to Week 26 for FBG.

5.12.1. Primary Analysis with Graphical Approach

The primary population for the key secondary efficacy analyses is the ITT population. Both the treatment regimen estimand and the efficacy estimand will be assessed using the same graphical testing scheme described in Section 5.3. Each set will use the overall 2-sided alpha of 0.05 to control overall Type 1 error rate after the primary objective has been achieved. The primary analysis is for the treatment regimen estimand for FDA, efficacy estimand by removing all metformin naïve patients for EMA, and efficacy estimand for all other regulatory agencies and disclosures. The same scheme will be adopted for each estimand, and each will be tested at the full significance level of 0.05. The graphical testing scheme is listed in Section 5.3.

All raw p-values generated from the following statistical models for the comparisons between All Dula or each dulaglutide dose and placebo at Week 26 will be used as the input data in the aforementioned graphical approach.

Change in HbA1c for each Dula arm (no pooling) and Placebo will be analyzed with the same model as defined in Section 5.11.1.

For the efficacy estimand, the MMRM model mentioned in Section 5.2 will be applied to change from baseline in BMI and FBG.

A longitudinal logistic regression model will be fitted to evaluate the proportion of patients achieving an HbA1c <7.0% at Week 26 using the same terms as described for the primary endpoint analysis by making these changes: adding logit as a link function in PROC GLIMMIX.

For the treatment regimen estimand, an ANCOVA model will be fitted to the complete data for change from baseline in BMI and FBG. This model includes stratification factors and treatment as fixed effects, and the corresponding baseline measure as a covariate. The missing data will be imputed (Section 5.11.3). A logistic regression model will be fitted to the complete data to analyze the proportions of patients reaching HbA1c target <7.0%. Stratification factors (insulin usage [yes, no], metformin usage [yes, no]) will be included as fixed effects, and baseline HbA1c as a covariate. Missing data at the endpoint will be imputed as not achieving the target.

5.12.2. Sensitivity Analyses

A logistic regression as a sensitivity analysis will be used to test the proportion of patients reaching HbA1c target of <7.0% at Week 26 in the ITT population by removing visit and visit-by-treatment interaction terms from the corresponding longitudinal logistic regression model described in Section 5.12.1. The patients who have been rescued or have missing data at Week 26 will be considered (imputed) as not having achieved the target.

5.13. Other Secondary Efficacy Analyses

The ITT population without postrescue therapy data (efficacy estimand) with all postfirst-dosing scheduled visits up to Week 52 (Visit 16) will be used for other secondary efficacy analyses.

These secondary endpoints include change from baseline to Week 52 in HbA1c; change from baseline to Week 52 in BMI and FBG; and the proportion of patients reaching HbA1c target of <7.0% at Week 52.

Each endpoint will be summarized by treatment group (Placebo/Dula 0.75, Dula 0.75, Dula 1.5, and All Dula) at each postdose visit. No inferential statistics will be generated. Within each treatment group, Week 26 and 52 corresponding responses will be provided using the corresponding statistical model described in Sections 5.12 and 5.11.1, respectively.

5.14. Pharmacokinetic/Pharmacodynamic Analyses

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

5.15. Safety Analyses

The safety parameters will include deaths, AEs, TEAEs, SAEs, AEs of special interest (AESI; which include pancreatitis, thyroid C-cell hyperplasia and C-cell neoplasms, and allergic/hypersensitivity reactions), hypoglycemic episodes, vital signs, ECG parameters, laboratory analytes including immunogenicity and hormones, and pubertal progression.

Unless otherwise stated, the ITT population will be used for analyses of the safety measurements and the all randomized population for data listings.

5.15.1. Adverse Events

An AE will be coded 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).

Summary statistics will be provided for the following: medical history including preexisting conditions, deaths, AEs, TEAEs, SAEs, AESIs, and study discontinuations due to death or AEs from baseline to Week 26 and from baseline to Week 52. The number of patients and proportions of patients experiencing the AEs will be reported for each treatment group, and Fisher's exact tests will be used to compare the treatment groups.

Treatment-emergent AEs are defined as events that are newly reported after the first dosing or reported to worsen in severity from the first dosing. If the first dosing date is missing, the randomized date will be used to impute the first dosing date for TEAE determination. The incidence of patients with at least 1 TEAE will be reported, and the proportion of patients experiencing each reported TEAE will be presented by PT, SOC, and treatment group. The proportion of patients experiencing each reported TEAE that are assessed as possibly related to the study treatment will also be summarized (see details in Section 5.2). Additionally, a summary of TEAEs by maximum severity will be presented descriptively by treatment group.

A listing of patients discontinuing the study due to AE or death will be generated.

5.15.2. Adverse Event of Interest

5.15.2.1. Hypoglycemia

Hypoglycemia will be classified as follows (ADA 2005; IHSG 2017):

- Documented Symptomatic Hypoglycemia: any time a patient feels that he or she is experiencing symptoms and/or signs associated with hypoglycemia, and has a plasma glucose (PG) level of <70 mg/dL (3.9 mmol/L) or PG <54 mg/dL (3 mmol/L). Therefore, the categories of interest will be as follows:
    - Documented Symptomatic Hypoglycemia with PG <70 mg/dL
    - Documented Symptomatic Hypoglycemia with PG <54 mg/dL
- Asymptomatic Hypoglycemia: an event not accompanied by typical symptoms of hypoglycemia, but with measured plasma glucose of <70 or <54 mg/dL. Therefore, the categories of interest will be as follows:
    - Asymptomatic Hypoglycemia with PG <70 mg/dL
    - Asymptomatic Hypoglycemia with PG <54 mg/dL
- Probable Symptomatic Hypoglycemia: an event during which symptoms of hypoglycemia are not accompanied by a PG determination (but that was presumably caused by a PG concentration of <70 or <54 mg/dL).
- Severe Hypoglycemia: an episode requiring the assistance of another person to actively administer carbohydrate, glucagon, or other resuscitative actions. These episodes may be associated with sufficient neuroglycopenia to induce seizures or coma. Plasma glucose measurements may not be available during such an event, but neurological recovery attributable to the restoration of PG to normal is considered sufficient evidence that the event was induced by a low PG concentration.
- Relative Hypoglycemia: defined as symptomatic events during which the person reports any of the typical symptoms of hypoglycemia, and interprets those as indicative of hypoglycemia, but with a measured PG concentration of >70 mg/dL (3.9 mmol/L).

- Nocturnal Hypoglycemia: any hypoglycemic event that occurs between bedtime and waking. Therefore, the categories of interest will be as follows:
    - Nocturnal Hypoglycemia with PG <70 mg/dL
    - Nocturnal Hypoglycemia with PG <54 mg/dL

Total hypoglycemia includes any event that meets criteria for documented symptomatic, severe, asymptomatic, probable, or unspecified (i.e., missing symptoms or signs, but with a PG <70 or <54 mg/dL) whether day-time or nocturnal. Those categories that are defined by PG will be analyzed by the PG <70 and <54 mg/dL thresholds separately. Cases of relative hypoglycemia will also be collected but will not be included in the category of total hypoglycemia.

Therefore, the total hypoglycemia category will have the following 2 subcategories:

1. Total Hypoglycemia (events with PG <70 mg/dL included)
2. Total Hypoglycemia (events with PG <54 mg/dL included)

The ITT population without postrescue data will be used as the population for the primary analysis for hypoglycemia.

Incidences and rates for documented symptomatic, severe, and total hypoglycemia will be summarized for a 1-year (365-day) period.

Additionally, a separate analysis will report all hypoglycemia events with PG <54 mg/dL, regardless of symptoms, and all severe hypoglycemia episodes for the ITT population, including events that occurred after patients started rescue therapy. The incidence of hypoglycemic episodes will be summarized for each treatment group for 0 to 26 weeks, 26 to 52 weeks, and 0 to 52 weeks. The summary includes the number of patients and percent of patients reporting hypoglycemic episodes for a given treatment at a certain visit interval. Fisher's exact test will be used for treatment comparison.

The rate for each category of the abovementioned hypoglycemia (PG <70 and PG <54 mg/dL if applicable) in episodes/patient/year for 0 to 26 weeks, 26 to 52 weeks, and 0 to 52 weeks will be analyzed using a negative binomial model via PROC GLIMMIX with log link function. The response is the frequency of each category of hypoglycemia of interest. The model will include stratification factors, treatment, and the interaction term of treatment by baseline insulin use as fixed effects. The logarithm of days between visits will be adjusted as an offset to account for possible unequal duration between visits and between patients. The predicted hypoglycemia rate per 1-year by treatment and visit interval will also be presented. The results will also be reported for each of the subgroups of baseline insulin use (yes, no).

The stratified Wilcoxon rank sum test will be applied for treatment comparison for all postdose intervals of interest. It will only be conducted and used as the primary analysis if the negative binomial regression analysis fails to converge.

Hypoglycemic episodes that occurred while patients are on/off insulin therapy will be summarized for the 3 categories of total hypoglycemia, documented symptomatic hypoglycemia,

and any hypoglycemia with PG <54 mg/dL with or without symptoms. A hypoglycemic event is considered to have occurred on insulin therapy if the event occurred within 24 hours after insulin use. Descriptive statistics for each treatment group will include the number of patients with events, the number of events, the total person years of exposure, and the number of events per 1000-person years of exposure. For each of the 3 categories of hypoglycemia, the proportions of patients with events while on insulin therapy as well as while not on insulin therapy will be calculated out of the total number of patients with events.

A listing of the individual hypoglycemic episodes will be presented using the all randomized population.

Additional exploratory analyses may be performed if deemed necessary.

5.15.2.2. Pancreatitis

The incidence of acute or chronic pancreatitis will be summarized descriptively for each treatment group separately for investigator-reported and adjudicated events from baseline to Week 26, from baseline to Week 52. The individual patients with this event will be listed along with lab tests, including lipase, amylase, and p-amylase.

5.15.2.3. Thyroid C-Cell Hyperplasia and C-Cell Neoplasms

The incidence of all types of thyroid tumors/neoplasms (benign and malignant) including C-cell hyperplasia will be summarized by each treatment group from baseline to Week 26, baseline to Week 52. The incidence of TEAEs associated with all types of thyroid tumors/neoplasms including C-cell hyperplasia will be reported for each treatment group from baseline to Week 26, baseline to V801. The individual patients with this event will be listed along with the laboratory tests, including calcitonin.

For calcitonin, summaries and analysis for changes from baseline and treatment emergent abnormal values will be provided as well as listing of abnormal values.

5.15.2.4. Nausea and Vomiting

Nausea and vomiting are the most common AEs reported in patients treated with dulaglutide from baseline to Week 26 and from baseline to Week 52. Summaries for onset (first incidence), duration, and severity of nausea and vomiting will be provided descriptively. Incidence of nausea and vomiting will be summarized using Fisher's exact test.

5.15.2.5. Allergic, Hypersensitivity, and Injection Site Reactions

Patients experiencing allergic, hypersensitivity, and injection site reactions will be listed and summarized for each treatment group from baseline to Week 26 and from baseline to Week 52 using Fisher's exact test.

5.15.2.6. Renal Impairment

The incidence of investigator-reported renal impairment (including renal failure) will be summarized for each treatment group from baseline to Week 26 and baseline to Week 52 by MedDRA terms.

5.15.3. Anti-Drug Antibodies (Immunogenicity)

If dulaglutide ADAs are detected, listings of antibody titers, antibody types, and HbA1c values by patient will be provided. A summary of incidence of treatment-emergent antibodies by antibody types will also be presented from baseline to Week 26 and baseline to Week 52. The presence of treatment-emergent ADAs is defined as a change from negative at baseline to positive at endpoint with antibody titer greater or equal to 1:4 or a positive at baseline to a positive at endpoint with greater or equal to a 4-fold increase. That is, if a positive antibody titer changes from 1:2 at baseline to 1:8 at endpoint, it is considered treatment emergent. Treatment-emergent ADAs may be further characterized and/or evaluated for their ability to neutralize the activity of dulaglutide.

5.15.4. Malignant Events

The incidence of investigator-reported malignant events will be summarized for each treatment group from baseline to Week 26 and baseline to Week 52 by PT.

5.15.5. Hyperglycemia

The number of patients and the percentage of patients who were rescued due to severe, persistent hyperglycemia will be summarized from baseline to Week 26 and from baseline to Week 52. Fisher's exact test will be applied to compare treatments on the proportion of patients who received rescue therapy due to severe, persistent hyperglycemia. Time to receive the first rescue therapy treatment due to severe, persistent hyperglycemia will be analyzed between treatment groups using the proportional hazard Cox regression model with treatment, stratification factors as fixed effects by excluding HbA1c strata, and adding baseline HbA1c as a covariate. A Kaplan-Meier curve will be plotted for each treatment group on time to first rescue therapy along with the log-rank p-value.

Patients who experienced severe, persistent hyperglycemia and received rescue therapy will be listed.

5.15.6. Vital Signs

All vital signs will be listed using the randomized population.

The change from baseline for sitting systolic and diastolic blood pressures and heart rate (HR)

(Table GBGC.5.2) will be analyzed using an MMRM model as described in Section 5.2 using all available post first-dosing scheduled visit data up to Week 52. There will be no multiplicity adjustments for analyses of vital signs. Corresponding figures may be presented.

The incidence of patients with select abnormal vital sign measurements will be summarized from baseline to Week 26 and from baseline to Week 52.

Table GBGC.5.2. Thresholds for Determining Abnormal Systolic Blood Pressure, Diastolic Blood Pressure, and Pulse Rate

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

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

5.15.7. Electrocardiograms

A listing of the ECG measurements (HR, PR, RR, QRS, QT, and corrected QT [QTc]), and abnormal selected ECG parameters will be produced in all randomized population separately.

Descriptive statistics for the actual measurements and change from baseline, by treatment arm and scheduled visit (9 and 16), will be performed for selected ECG parameters using ITT population. The parameters that will be included in the summary and analysis with MMRM model described in Section 5.2 are HR, RR, PR, and QRS intervals.

QT Fridericia's and Bazett's corrections will be used to correct the QT interval using the formulas below:

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

$$QTcB = QT/RR^{1/2}$$

A MMRM for QT, QTcF, QTcB change from baseline will be analyzed using MMRM. In this model, RR change from baseline and corresponding baseline QT will be fitted as covariates with treatment, insulin use ("yes" versus "no"), metformin use ("yes" versus "no"), baseline HbA1 as strata, visit, and treatment-by-visit as fixed effects. All analysis results related to QTcB are not intended for the CSR but for publication only.

For quantitative ECG variables, summaries and analyses of treatment-emergent abnormal values, and of medical query: any term in arrhythmia-related investigations and signs and symptoms will be generated separately.

Selected thresholds for HR, PR interval, and QTc interval are shown in Table GBGC.5.3 and will be summarized clinically relevant abnormal values for these variables from baseline to Week 26 and from baseline to Week 52.

Table GBGC.5.3. Thresholds for HR, PR Interval, and QTc Interval

ECG Variable (unit)	Threshold
QTcB, QTcF actual measurement (msec)	>460 (≤15 yrs), >450 (male, ≥16 yrs), >470 (female, ≥16 yrs) >500
QTcB, QTcF change from Baseline (msec)	>30, >60, >75
ECG heart rate (bpm)	Refer to pulse rate in Table GBGC.5.2
PR interval (msec)	<120 ≥220

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; HR = heart rate; yrs = years.

5.15.8. Analysis of Laboratory Analytes

All laboratory measurements (including scheduled and unscheduled) will be listed using all randomized population. A listing of patients with treatment-emergent abnormal laboratory, with the normal ranges and potentially clinically relevant reference limits for all laboratory tests, with serum calcitonin levels suggestive of thyroid C-cell abnormalities, will be listed separately. For certain labs, age-dependent reference ranges will be used as appropriate.

Descriptive summary statistics will be presented for sodium, potassium, calcium, and glucose random at each scheduled visit in raw data and change from baseline without formal statistical analysis.

For each continuous laboratory measurement listed below, the change from baseline will be summarized for each treatment. The summary statistics include number of patients, mean, SD, minimum, Q1, median, Q3, and maximum. These analytes will be analyzed using ANOVA on the ranks, with treatment as a fixed effect. LOCF will be used to impute missing postbaseline values. Wilcoxon signed-rank test will be used to compare baseline and each post-baseline visit within the same treatment:

- Pancreatic enzymes (lipase, amylase, P-amylase)
- Liver test panel (ALT, AST, CK, alkaline phosphatase, total bilirubin, direct bilirubin, albumin)
- Renal (eGFR, albumin-to-creatinine ratio [ACR], serum creatinine, serum BUN, serum cystatin C, uric acid)
- Serum lipids (total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides)
- Hormone-related safety assessed by the effect on serum
    - Prolactin
    - Insulin-like growth factor-1

  - Testosterone (males only)
  - Cortisol
  - Estradiol
  - Luteinizing hormone
  - Calcitonin
- hematology (hemoglobin, hematocrit, RBC, mean cell volume, mean cell hemoglobin concentration, WBC, neutrophils [segmented], lymphocytes, monocytes, eosinophils, basophils, platelets)

The above lab measures will also be compared to reference range to determine whether they are abnormally high, low, or normal. The incidence and percent of high, low, and normal values will be summarized for each of the treatment arms and compared using the Fisher's exact test.

Shift tables of the change from maximum baseline to maximum postbaseline, from baseline to Week 26, and from baseline to Week 52 in the selected analytes using clinical meaningful thresholds will be summarized separately. The selected analytes include creatinine, ALT, AST, ACR, and calcitonin.

Similar analysis will be performed on the change from minimum baseline to minimum post-baseline, and the change from baseline to Week 26, and from baseline to Week 52 in eGFR, creatinine, ALT, AST, ACR, and calcitonin to that for the change from maximum baseline to maximum postbaseline.

Summary of pancreatic enzymes (p-amylase, amylase, lipase) by threshold (>1×upper limit of normal [ULN], ≥3×ULN) from screening to Week 52 will be summarized descriptively using all ITT population, ITT population with normal range baseline, ITT population with >1×ULN baseline.

The number of patients and percentage of patients with a serum calcitonin ≥35 pg/mL and a simultaneous increase of 50% from baseline, and patients with a serum calcitonin ≥20 pg/mL and <35 pg/mL with a simultaneous increase 50% from baseline to Week 52 will be summarized using Fisher's exact test.

A summary of the number of patients and percentage of patients exceeding 2×, 3×, 5×, and 8×ULN for each treatment at baseline, Week 26, and Week 52 will be presented for ALT, AST, and total bilirubin using Fisher's exact test.

A summary of the number of patients and percentage of patients meeting notable criteria for potential hepatotoxicity (ALT or AST >3×ULN and total bilirubin >2×ULN) for each treatment group at each postdose visit will be presented.

The number of patients and percentages for patients with pancreatic enzymes above ULN and greater than or equal to 3×ULN will be summarized at baseline and by visit for each treatment group.

Number of patients and percentage for female patients showing positive outcome from serum pregnancy test will be summarized descriptively for each treatment group from baseline to Week 26, from Week 26 to Week 52, and from baseline to Week 52.

Number of patients and percentage of subjects of normal and abnormal values for each urinalysis test: protein, glucose, ketones, blood, and leukocyte esterase will be presented descriptively for each treatment group from baseline to Week 26 and from baseline to Week 52.

Additional analyses may be conducted if deemed necessary.

5.15.9. Pubertal Progression Evaluation

A shift table will be presented in those pubertal patients at baseline (Visit 2) to evaluate the change in Tanner Staging from baseline to Week 26, and baseline to Week 52 separately for male and female group at each treatment group without pooling. This includes no change and an increase in 1, 2, 3, and 4 levels. The likelihood ratio p-value will be the test statistic.

Morning serum hormone (estradiol, testosterone, LH, IGF-1, cortisol, and prolactin) will be analyzed with the method described in Section 5.15.8.

5.16. Exploratory Analysis

The population for all exploratory analyses is the ITT population, which treats all postrescue therapy measurements as missing (efficacy estimand) with all scheduled postdose visit data up to Week 52 (Visit 16) unless otherwise stated.

5.16.1. Parameters of Interest

The exploratory parameters of interest include, but are not limited to, the following measurements in change from baseline to Week 26 and 52 (unless otherwise specified):

- HbA1c to Week 13 (Visit 6)
- Percentage of patients reaching HbA1c target of:
  - ≤6.5%
  - ≤6.5% without severe, documented symptomatic (blood glucose [BG] <70 mg/dL), or probable hypoglycemic episodes
  - <7.0% without severe, documented symptomatic (BG<70 mg/dL), or probable hypoglycemic episodes
- Body weight
- BMI standard deviation score (SDS)
- Height and height SDS
- Waist circumference
- Measures of insulin resistance, beta cell function, and serum adiponectin at Weeks 13 and 26

- Change in basal insulin dose from baseline to Week 26 and from baseline to Week 52

5.16.2. Analysis on HbA1c Change from Baseline to Week 13

The same MMRM model as that described in Section 5.11.1 will be used. The Week 13 data will be included in the analysis of the data through Week 26 and through Week 52.

5.16.3. Analysis of Percentage of Patients Reaching HbA1c Target

The same model as that described in Section 5.12.1 for the percentage of patients reaching a HbA1c target of <7% will be applied for 3 parameters of interests: HbA1c ≤6.5%, composite endpoint of HbA1c <7% without severe, documented symptomatic (BG <70 mg/dL), or probable hypoglycemic episodes, and HbA1c ≤6.5% without severe, documented symptomatic (BG <70 mg/dL), or probable hypoglycemic episodes. Detailed hypoglycemia episodes are defined in Section 5.15.2.1.

Additional analysis will be performed on the ITT population using logistic regression defined in Section 5.12.2 for Week 26 and Week 52, where patients who have been rescued or have missing data at Week 26 or Week 52 will be imputed to have not achieved the target, while for nonrescued patients the available data will be used.

5.16.4. Analysis of Change in Body Weight, BMI SDS, Height and Waist Circumference

The standard MMRM approach detailed in Section 5.2 for analyzing continuous data will be used to analyze change from baseline in body weight, BMI SDS, height, height SDS, and waist circumference. When analyzing height, sex will also be included in the model as a covariate since this is a prognostic factor for height in children and adolescents. For BMI and height, SDS will be calculated for each subject based on age and sex using World Health Organization standards and methods (WHO 2008) and analyzed as described above.

5.16.5. Analysis of Change in Basal Insulin Dose

Descriptive summary statistics will be presented for the change from baseline in basal insulin dose in both U/Day and U/kg/Day for each treatment at each visit from baseline to Week 26 and from baseline to Week 52.

5.16.6. Pharmacodynamic Endpoints on Insulin Sensitivity and β-Cell Function

The endpoints include HOMA2-%B, HOMA2-IR, 1/fasting insulin or 1/fasting C-peptide, and adiponectin. HOMA1-%B is used in Lilly original proposal to FDA. However, HOMA1 as the original model developed in 1985 did not account for differences between hepatic and peripheral insulin sensitivity, increases in insulin secretion, or decreases in hepatic glucose production for plasma glucose concentrations above 180 mg/dL, renal glucose losses, or the contribution of circulating proinsulin. In addition, the original HOMA model (HOMA1) uses equations that were calibrated to insulin assays used in the 1970s, which result in underestimation of %-S and

overestimation of %-B. An updated HOMA model (HOMA2) has been created to account for these variations. With knowledge of these differences, it is therefore important for this study to use HOMA2 instead of HOMA1 to quantify insulin resistance and beta-cell function.

The estimated insulin sensitivity scores will be calculated as follows: $\log_e IS = 4.64725 - 0.02032$ (waist, cm) – 0.09779 (HbA1c, %) – 0.00235 (TG; mg/dl).

The normality assumption of each biomarker will be assessed both visually (histogram and QQ-plot) and through the Shapiro-Wilk test prior to the formal statistical analysis.

The MMRM model will be fitted for each biomarker separately with stratification factors, treatment, time, visit-by-time interaction as fixed effect, subject as a random effect, and corresponding baseline as a covariate. If data are normal distributed, change from baseline will be the parameter of interest. If data are log normal distributed, log (ratio relative to baseline) will be the parameter of interest.

5.17. Subgroup Analyses

A subgroup analysis will be performed on the primary endpoint using the treatment regimen estimand with the washout method for multiple imputation for missing data, and the same model as the primary analysis model by adding 3-way interaction of visit-by-treatment-by-subgroup, 2-way interaction of treatment-by-subgroup, and visit-by-subgroup. The 2-way interaction of treatment-by-subgroup at the primary time point of 26 weeks will be evaluated to assess an interaction in the treatment effect with the subgroup levels. Significance will be evaluated at 2-sided alpha of 0.1. The following are candidate subgroups that might be analyzed. This list is not necessarily all-inclusive:

- gender
- age group (≤14 years old, >14 years old)
- race
- ethnicity
- region (US and non-US)
- duration of diabetes at baseline (<median duration and ≥median duration)
- baseline BMI (<median and ≥median)
- baseline body weight (<median and ≥median)
- metformin usage (“yes” or “no”)
- baseline HbA1c (≤8.0%, >8.0%)
- basal insulin usage (“yes” or “no”)
- monotherapy (“yes” or “no”)
- metformin and insulin usage (“yes” or “no”)

A descriptive summary of the change from baseline to Week 26 in HbA1c, FBG, and BMI, and HbA1c target of <7% at Week 26 will be presented by age group, categorized by (10 to 14 years) and (>14 years).

A descriptive summary for the key clinically important safety endpoints may be presented for the aforementioned age group.

5.18. Patient-Reported Outcome Analyses

All PRO analyses are exploratory. Demographic and baseline characteristics of the pediatric population at baseline for HO population will be summarized. The EQ-5D-Y is a standardized generic measure of health status developed by the EuroQol Group. The EQ-5D-Y consists of the EQ-5D-Y descriptive system and the EQ VAS. The descriptive system comprises the same 5 dimensions as the EQ-5D 3 level (EQ-5D-3L), but using a child-friendly wording (mobility; looking after myself; doing usual activities; having pain or discomfort; and feeling worried, sad, or unhappy). Each dimension has 3 levels: no problems, some problems, or a lot of problems. The EQ VAS records the respondent's self-rated health on a vertical VAS where the endpoints are labelled "The best health you can imagine" and "The worst health you can imagine." This VAS information can be used as a quantitative measure of the perception of their overall health by the individual respondents.

The EQ-5D-Y health states, defined by the EQ-5D descriptive system, will be converted into a single index value by applying a formula that essentially attaches value (also called weights) to each of the levels in each dimension. However, as of 2020, an EQ-5D-Y index formula is not available for converting the health states into a single index value. The analysis for the PRO measures (EQ-5D-Y VAS and EQ-5D-Y index scores only if available) will be MMRM for the change in PRO score from baseline to each postdose visit for the HO population. The MMRM will include treatment, insulin usage, metformin usage, baseline HbA1c stratification, visit, visit-by-treatment interaction as fixed effects, and baseline score as a covariate. The variance-covariance structure for this model will be selected in the same order defined in Section 5.2. The EQ-5D-Y index score will be analyzed if it is available. Patient responses by EQ-5D-Y domains (mobility; looking after myself; doing usual activities; having pain or discomfort; feeling worried, sad, or unhappy) will also be descriptively examined. The above analyses will be calculated and reported for Age <13 years old, ≥13 years old, and ≤14 years old, >14 years old, and all age combined group. Please refer to the EQ-5D-Y User's Manual for appropriate descriptive reporting examples, VAS, and how to handle missing data.

The 23-item PedsQL Generic Core Scales were designed to measure the core dimensions of health as delineated by the World Health Organization, as well as role (school) functioning. The 2 versions used are for children (ages 8 to 12) and for teens (ages 13 to 18). The generic scale comprises 4 multidimensional scales and 2 Summary Scores along with scores calculated for each of the 4 dimensions and a total score. The 4 dimensional scales are Physical Functioning (8 items), Emotional Functioning (5 items), Social Functioning (5 items), School Functioning (5 items), and the 2 Summary Scores, which are Psychosocial Health Summary Score (15 items),

and Physical Health Summary Score (8 items), and a Total Score (23 items). Higher scores indicate better health-related quality of life (HRQOL).

The PEDS-QL 3.2 Diabetes Module is a diabetes-specific HRQOL measurement instrument that includes 33 items comprising 5 dimensions for ages 13 to 45 years. For ages 2 to 12 years, the PedsQL 3.2 Diabetes Module is composed of 32 items comprising 5 dimensions (1 less item for the Worry Scale). The 5 dimensions of this diabetes specific module are Diabetes (15 items), Treatment I (5 items), Treatment II (6 items), Worry 2 items (3 items for teens and adults), and Communication (4 items). Scores are calculated for each of the 5 dimensions and a total score is available. Higher scores indicate lower problems.

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

The PEDS-QL Diabetes [3.2] Worry Dimension will be analyzed via an MMRM for the change in score from baseline to Week 26 and Week 52 in HO population. The model will include treatment, insulin usage, metformin usage, baseline HbA1c stratification, visit, and visit-by-treatment interaction as fixed effects and baseline score as a covariate. The variance-covariance structure for this model will be selected in the same order defined in Section 5.2. Baseline data and each postdose visit scores for the Worry Dimension will be calculated and reported for Age <13 years old, ≥13 years old, and ≤14 years old, >14 years old, and all age combined group.

The analysis for the PRO measures (PEDS-QL Diabetes [3.2] [except Worry Dimension as it is already specified above] and PEDS-QL [generic scale]) will use similar MMRM model to that for PEDS-QL worry dimension score. The analyses will cover the total scores, all subscale scores, and dimensions. Baseline data and each visit scores will be calculated and reported for Age <13 years old, ≥13 years old, and ≤14 years old, >14 years old, and all age combined group.

Details about scoring the PRO instruments can be found in the Peds-QL Scoring Manual (Mapi Research Trust 2015 [WWW]) and EQ-5D-Y User Guide (EuroQol Research Foundation 2015 [WWW]).

6. Unblinding Plan

6.1. Interim Analyses

An independent DMC will have the responsibility to review the interim analysis results in order to monitor the safety of the patients in the study until the last patient reaches the primary endpoint at 26 weeks. The detailed analysis and communication plan for the interim analyses will be defined in a separate DMC charter. An internal SAC but external to the study team will perform the data analysis for the DMC. As no efficacy analyses are planned by the DMC, the Family-wise Error Rate (FWER) will not be affected by any of these interim analyses; hence no alpha spending is necessary.

6.1.1. Safety Analysis

The DMC will conduct a safety interim approximately every 6 months with initiation after 50 patients are enrolled into the study, as deemed appropriate based on the observed enrollment rate, until the last patient reaches the primary endpoint at 26 weeks. However, the number of interims will be determined by the enrollment rate: if enrollment is much slower than expected, safety interims may begin sooner and may occur less frequently (annually).

6.1.2. Sample Size Re-Estimation

As described in Section 5.1, a sample size re-estimation calculation may be conducted by the SAC. Based on this calculation, the SAC will recommend to the sponsor either increasing the total enrollment or making no changes to the target enrollment based on pre-specified rules. Only the SAC will be unblinded for this analysis. The only information provided to the sponsor by the SAC will be the recommended sample size.

For further details on the statistical methodology, see Section 5.1.

6.1.3. Pharmacokinetic Analysis

A limited number of preidentified individuals external to the study team may gain access to the unblinded data during the double-blind phase prior to the primary database lock at 26 weeks in order to initiate the population PK/PD model development processes. These population PK and PK/PD models may be refreshed with data after final database lock or evaluated graphically. Information that may unblind the study during the analyses will not be reported to study sites or to the blinded study team until the study has been unblinded.

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 Week 26 (Visit 9). The treatment assignments will be blinded to patients and investigators until the end of Visit 9.

Emergency unblinding 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 unblinding event are recorded and reported by the IWRS.

The investigator should make every effort to contact the Lilly clinical research physician (CRP) or designee prior to unblinding a patient's treatment assignment. If a patient's treatment assignment is unblinded, Lilly must be notified immediately.

If an investigator, site personnel performing assessments, or patient is unblinded during the double blinded period, the patient will continue in the study (Protocol Section 9.7).

6.3. Sponsor/Trial Level Unblinding

The study team will remain blinded to treatment assignments until all patients have completed Week 26 (Visit 9) and the database has been locked for primary database lock.

7. References

[ADA] American Diabetes Association; Workgroup on Hypoglycemia. Defining and reporting hypoglycemia in diabetes: a report from the American Diabetes Association Workgroup on Hypoglycemia. *Diabetes Care*. 2005;28(5):1245-1249.

[ADA] American Diabetes Association. Glycemic targets: standards of medical care in diabetes-2020. *Diabetes Care* 2020;43(suppl 1):S66-S76.

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

EuroQol Research Foundation. 2015. Available at: www.euroqol.org. Accessed June 4, 2015.

Fu H, Manner D. Bayesian adaptive dose-finding studies with delayed responses. *J Biopharm Stat*. 2010;20(5):1055-1070.

[IHSG] International Hypoglycaemia Study Group. Glucose concentrations of less than 3.0 mmol/L (54 mg/dL) should be reported in clinical trials: a joint position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care.* 2017;40(1):155-157.

Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H. The prevention and treatment of missing data in clinical trials. *N Engl J Med.* 2012;367(14):1355-1360.

Mapi Research Trust. 2015. PROQOLID. Available at: http://www.proqolid.org. Accessed June 4, 2015.

Mehta CR, Tsiatis AA. Flexible sample size considerations using information-based interim monitoring. *Drug Information Journal*. 2001;35(4):1095-1112.

[NRC] National Research Council of the National Academies. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: *The National Academies Press*, 2010.

O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clin Pharmacol Ther*. 2012;91(3):550-554. doi:10.1038/clpt.2011.340.

Permutt T. Sensitivity analysis for missing data in regulatory submissions. *Stat Med*. 2016;35(17):2876-2879.

Pritchett Y, Jemiai Y, Chang Y, Bhan I, Agarwal R, Zoccali C, Wanner C, Lloyd-Jones D, Cannata-Andía JB, Thompson T, Appelbaum E, Audhya P, Andress D, Zhang W, Solomon S, Manning WJ, Thadhani R. The use of group sequential, information-based sample size re-estimation in the design of the PRIMO study of chronic kidney disease. *Clin Trials*. 2011;8(2):165-174.

[WHO] World Health Organization Multicentre Growth Reference Study Group. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization, 2006.

Leo Document ID = faf73dbd-f721-4918-9c2b-574750e018b9

Approver: PPD

Approval Date & Time: 18-Feb-2021 17:40:56 GMT

Signature meaning: Approved