H9X-JE-GBGF Statistical Analysis Plan

A Phase 4 Study of Efficacy and Safety of Dulaglutide When Added to Insulin Treatment With or Without Oral Antidiabetic Medication in Patients with Type 2 Diabetes

NCT027504

Approval Date: 12-Sep-2018

1. Statistical Analysis Plan for Clinical Studies: H9X-JE-GBGF: A Phase 4 Study of Efficacy and Safety of Dulaglutide When Added to Insulin Treatment With or Without Oral Antidiabetic Medication in Patients with Type 2 Diabetes

Dulaglutide (LY2189265) Type 2 Diabetes

Study H9X-JE-GBGF is a Phase 4, randomized, placebo-controlled, double-blind (16 weeks) and subsequent open-label (36 weeks) study to assess the efficacy and safety of dulaglutide 0.75 mg in Japanese patients with type 2 diabetes who have inadequate glycemic control on insulin therapy with or without 1 or 2 oral antidiabetic medications.

Eli Lilly Japan K.K.
Kobe, Hyogo Japan

Protocol H9X-JE-GBGF
Phase 4

Statistical Analysis Plan electronically signed and approved by Lilly on date provided below.

Approval Date: 12-Sep-2018 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Appendix **Page**

3. Revision History

SAP Version 1 was approved prior to First Patient Visit (FPV).

SAP Version 2: Unblinding Plan Section was added to clarify the unblinding plan and blinding strategy throughout the study conduct. Summary of concomitant medication and plot of SMPG during follow-up period have been added.

SAP Version 3: Some post hoc analyses were added: Statistical tests were added for DTSQs and DTR-QOL.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to show superiority of the addition of once-weekly dulaglutide 0.75 mg compared to the addition of once weekly placebo to insulin treatment with or without 1 or 2 OADs, on change from baseline in HbA1c after 16 weeks of treatment in Japanese patients with T2D.

4.2. Secondary Objectives

The secondary objectives of this study are as follows:

- For efficacy, to compare dulaglutide 0.75 mg to placebo at 16 weeks;
- For safety, to compare dulaglutide 0.75 mg to placebo for 16 weeks;
- To describe long-term efficacy of dulaglutide 0.75 mg in combination with insulin therapy up to 52 weeks;
- To describe long-term safety of dulaglutide 0.75 mg in combination with insulin therapy up to 56 weeks (to include the follow-up period).

4.3. Exploratory Objectives

The exploratory objective of this study is as follows:

- To assess the health outcomes variables at 16 and 52 weeks.

5. A Priori Statistical Methods

5.1. Determination of Sample Size

A total of approximately 160 patients (dulaglutide: 120 patients; placebo: 40 patients) will be randomized to have approximately 136 completers (dulaglutide : 102 patients; placebo: 34 patients) to show dulaglutide is superior to placebo with more than 99% power assuming a treatment difference of 1.0% in HbA1c reduction, standard deviation (SD) = 1.0%, and dropout rate of 15%. The screen failure rate is estimated as 20%. Approximately 200 patients will be screened.

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 clinical study report (CSR). Additional exploratory analyses of the data will be conducted as deemed appropriate.

All analyses will be implemented using Version 9.1.3 SAS or higher.

Continuous data will be summarized in terms of the number of observations (n), mean, standard deviation (SD) and/or standard error of the mean (SEM), median, minimum (Min), and maximum (Max).

Categorical data will be summarized in terms of the number of observations (n), frequency count, and percentages. Percentages will be rounded to one decimal place. Percentages will not be presented for zero counts. Unless noted otherwise, percentages will be calculated using n (the number of observations with non-missing values) as denominator.

For efficacy analysis based on mixed-model for repeated measures (MMRM), summary statistics will be provided for the observed value and change from baseline value for each arm based on MMRM. Also, for efficacy analysis, statistical tests (including MMRM and ANCOVA) will only be conducted during the double-blind primary treatment period at each visit (i.e. Visit 4, 5, 6 and 7) and will not be conducted for time point after Visit 7.

For safety analysis, summary statistics will be provided for each arm during the both double-blind primary treatment period (i.e. Visit 3 to 7) and entire study period (i.e. Visit 3 to 13). Table GBGF 5.1 gives the treatment groups to be displayed for each treatment period.

Unless otherwise specified, 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. There will be no multiplicity adjustment for pairwise treatment comparisons.

List of analysis (LOA) will contain the list of all tables, figures and listings (TFLs), and will be provided separately.

Table GBGF 5.1. Treatment Groups for Each Treatment Period

Treatment Period	Treatment Groups	Abbreviation	Between-Group Comparison When Applicable
Double-blind primary treatment period (i.e. Visit 3 to 7)	Placebo Dulaglutide Total	Plc Dula Total	Dula vs Plc
Treatment period (i.e. Visit 3 to 12)	Placebo/ Dulaglutide Dulaglutide/ Dulaglutide Total	Plc/Dula Dula/Dula Total	No Between-Group Comparisons
Entire study period (i.e. Visit 3 to 13)	Placebo/ Dulaglutide Dulaglutide/ Dulaglutide Total	Plc/Dula Dula/Dula Total	No Between-Group Comparisons

5.2.1. Analysis Populations

The analysis populations used in the study are described below (Table GBGF 5.2).

Table GBGF 5.2. Analysis Population for Study H9X-JE-GBGF

Population	Definition
All Entered	All patients who signed informed consents
All Randomized	All patients who were randomized to a treatment arm
Non-randomized	All patients entered but not randomized to a treatment arm
Full Analysis Set (FAS)	All patients who received at least one dose of study treatment and have at least one measurement of HbA1c after study treatment
Per-Protocol Set (PPS)	All patients in the FAS who meet the following criteria: • Have no important protocol deviations that could impact the assessment of the primary objective (see section 5.4) • At least 75% compliant with the required doses of investigational product • Complete the treatment period (16 weeks [Visit 7]) for primary endpoint • Not missing more than 2 consecutive weekly injections of investigational product or more than 14 consecutive daily injections of insulin
Safety Analysis Set	All patients who received at least one dose of study treatment

5.2.2. Baseline Definition

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

Change from baseline will be calculated as the post-baseline value minus the baseline value. If all baseline values are missing for a particular variable, the change from baseline will not be calculated.

5.3. Patient Disposition

For summary tables, frequency counts and percentages of all patients entered, randomized, completed, and/or discontinued from the study will be presented, and the reasons for discontinuation from the study will be summarized. Listing will also be made.

5.4. Protocol deviations

Important protocol deviation will only be summarized and listed. For the protocol deviations resulting in exclusion of the patient from the PPS, see the Protocol Deviation Pre-defined List separately provided.

5.5. Patient Characteristics

Demographic (gender, age, etc.) and baseline variables (body weight, height, baseline HbA1c, etc.) will be summarized using both FAS and PPS. Listing will also be made using FAS.

5.6. Concomitant Therapy

Concomitant therapy and type of insulin will be summarized and listed using FAS. Oral antidiabetic medicine will be summarized separately. WHO drug generic name will be used. The definition of each type of insulin regimen is as follows.

- Basal insulin regimen is defined as the regimen using only basal insulin;
- Premixed insulin regimen is defined as the regimen using only premixed insulin;
- Basal/mealtime insulin regimen is defined as the regimen using only both basal insulin and mealtime insulin.

5.7. Historical Illness

Historical illnesses will be listed using all randomized patients. No summaries will be created for this measurement. Pre-existing condition will also be summarized and listed using safety analysis set.

5.8. Treatment Compliance

Overall treatment compliance, overall investigational product compliance, treatment compliance (not missing more than 2 consecutive weekly injections of investigational product) and treatment compliance (not missing >14 consecutive daily injections of insulin) will be summarized by treatment group using FAS. Listing will also be made using FAS.

The overall treatment compliance is defined as taking at least 75% of the required doses of investigational product and not missing more than 2 consecutive weekly injections of dulaglutide or >14 consecutive daily injections of insulin. Similarly, a patient will be considered non-compliant if he/she is judged by the investigator to have intentionally or repeatedly taken more than the prescribed amount of medication. The overall treatment compliance in percentage will be calculated by the number of patients who meet overall treatment compliance divided by the total number of patient *100.

In addition, the overall investigational product compliance will be calculated for each patient. This will be calculated by taking the number of injections the patient took divided by the total number of scheduled injections for this patient.

The treatment compliance (not missing more than 2 consecutive weekly injections of investigational product) in percentage will be calculated by the number of patients who meet treatment compliance (not missing more than 2 consecutive weekly injections of investigational product) divided by the total number of patient *100.

The treatment compliance (not missing >14 consecutive daily injections of insulin) in percentage will be calculated by the number of patients who meet treatment compliance (not missing >14 consecutive daily injections of insulin) divided by the total number of patient *100.

5.9. Treatment Exposure

Treatment exposure will be summarized and listed using safety analysis set.

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

Duration of exposure will be categorized into the following groups: >0 weeks, ≥16 weeks, and ≥52 weeks. These categories will be summarized as frequency by treatment.

5.10. Efficacy Analysis

5.10.1. Primary Efficacy Analysis

Analysis of HbA1c based on MMRM will be summarized by treatment group using FAS. The primary analysis model is shown in 5.10.1.2.

5.10.1.1. Additional Analyses for Primary Endpoint

Analysis of HbA1c based on MMRM will be summarized by treatment group using PPS. Also, analysis of HbA1c based on ANCOVA will be summarized by treatment group using FAS.

5.10.1.2. Primary Analysis Model

The primary analysis model will be an MMRM for HbA1c change from baseline to 16 weeks (Visit 7) using REML with treatment, insulin regimen (basal insulin, premixed insulin, or

basal/mealtime insulin), visit, and treatment-by-visit 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. 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 sum of squares.

5.10.1.3. Secondary Analysis Model

The secondary analysis for the primary endpoint will be an ANCOVA for HbA1c change from baseline to the Week 16 endpoint with treatment, and insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin) as fixed effects and baseline HbA1c as a covariate, with missing endpoints imputed with last observation carried forward (LOCF) using postbaseline data only. If there are no data after the date of randomization, the endpoint will be considered missing.

5.10.2. Secondary Efficacy Analysis

Secondary efficacy measures that are continuous variables (changes from baseline) other than HbA1c will be analyzed using MMRM on the FAS. The MMRM model will include treatment, baseline HbA1c (<8.5%, ≥8.5%), and insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin) as fixed effects and a covariate of baseline value. Also, the ANCOVA models will include treatment, baseline HbA1c (<8.5%, ≥8.5%), and insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin) as fixed effects and a covariate of baseline value.

5.10.2.1. Percentage of Patients Achieving HbA1c of <7.0% and ≤6.5%

Analysis of patients achieving HbA1c levels <7.0% or ≤6.5% based on repeated measures logistic regression with generalized linear mixed model will be summarized by treatment group using FAS. The model will include independent variables of treatment, insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin), visit, and treatment-by-visit interaction as fixed effects, and baseline HbA1c as a covariate.

5.10.2.2. Analysis of Fasting Serum Glucose (FSG)

Summary and analysis of FSG based on MMRM will be displayed by treatment group using FAS.

5.10.2.3. Analysis of 7-point SMPG profile

The 7-point SMPG profiles before and 2 hours after breakfast, before and 2 hours after lunch, before and 2 hours after dinner, and bedtime will be obtained within 2 weeks prior to corresponding visit. In general, the mean of the two daily mean values will be used for reporting purpose. If the value from either day is missing, then the non-missing value will be used as the mean of the two daily mean values. Also, if both two daily mean values are missing then the mean of the two daily mean values will be missing. The daily mean is calculated as the mean of available values collected on a particular day.

The following variables for 7-point SMPG profile will be analyzed using ANCOVA model with missing endpoints imputed with LOCF:

1. Pre morning meal PG (mg/dL)
2. 2-hour postprandial measurement for morning meal PG (mg/dL)
3. Pre midday meal PG (mg/dL)
4. 2-hour postprandial measurement for midday meal PG (mg/dL)
5. Pre evening meal PG (mg/dL)
6. 2-hour postprandial measurement for evening meals PG (mg/dL)
7. Bedtime BG (mg/dL)
8. Morning meal 2-hr excursion (mg/dL)
9. Midday meal 2-hr excursion (mg/dL)
10. Evening meal 2-hr excursion (mg/dL)
11. Mean of all meals 2-hr excursion
12. Mean of all 7-point PG (mg/dL)
13. Mean of all pre-meals PG (mg/dL)
14. Mean of all postprandial meals PG (mg/dL)
15. Circadian variation in 7-point PG (mg/dL)

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

For the morning, midday and evening meal 2-hr excursion (reports 8-10), the daily excursion for each meal is calculated as the postprandial meals PG minus the fasting plasma glucose (FPG). The mean glucose excursion for each meal at each visit is calculated as the average of 2 daily glucose excursions. The change from baseline is calculated as the mean excursion at endpoint minus the mean excursion at baseline.

For mean of all meals 2-hr excursion at each visit (report 11), 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-hr 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-hr excursion at endpoint minus the mean of all meals 2-hr excursion at baseline.

For mean of all 7-point PG (report 12), the daily mean is calculated as the average of 7 PG values collected on a particular day. The mean of all 7-point PG 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 PG at endpoint minus the mean of all 7-point PG at baseline.

For mean of all pre-meals PG (report 13), the pre-meal daily mean is calculated as the average PG values collected for pre-morning, pre-midday and pre-evening meals on a particular day. The mean of all pre-meals PG 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-meals PG at endpoint minus the mean of all pre-meals PG at baseline.

For mean of all postprandial meals PG (report 14), the post-meal daily mean is calculated as the average of 2-hr postprandial meals PG values of morning, midday and evening meals on a particular day. The mean of all postprandial meals PG 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 meals PG at endpoint minus the mean of all postprandial meals PG at baseline.

For circadian variation in 7-point PG (report 15), the daily circadian variation is calculated as the difference between maximum and minimum (maximum - minimum) PG values collected on a particular day. The mean circadian variation at each visit is calculated as the average of 2 daily circadian variation. The change from baseline is calculated as the circadian variation at endpoint minus the circadian variation at baseline.

Fasting PG during follow-up period will be plotted by patient.

5.10.2.4. Analysis of Body Weight

Analysis of body weight based on MMRM will be summarized by treatment group using FAS.

5.11. Insulin Dose

For the analysis of insulin dose, only descriptive statistics will be provided.

Daily insulin dose (unit and unit/kg body weight) for each insulin regimen (i.e. basal insulin, premixed insulin, or basal/mealtime insulin) at the baseline and specific time point (i.e. Visit 7 and Visit 12), and change from baseline will be summarized and listed using FAS.

For the patients using basal/mealtime insulin, daily insulin dose (unit and unit/kg body weight) for each insulin type (i.e. basal or mealtime insulin) at the baseline and specific time point (i.e. Visit 7 and Visit 12), and change from baseline will be summarized using FAS

5.12. Health Outcome Analyses

5.12.1. DTSQs (Diabetes Treatment Satisfaction Questionnaire status)

Summary statistics for each dimension and total score by visit and change from baseline values will be provided using FAS. Hyperglycemia, hypoglycemia, total score except blood glucose control score and total score at Week 16 will be compared between treatment groups using an ANCOVA models that includes treatment and insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin) as fixed effects and a covariate of baseline value. The LS means and 95% confidence intervals for treatment difference and p-values for comparisons will be reported.

5.12.2. DTR-QOL (Diabetes Therapy Related-Quality of Life)

Summary statistics for each dimension and total score by visit and change from baseline values will be provided using FAS. Each dimension and total score at Week 16 will be compared between treatment groups using an ANCOVA models that includes treatment and insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin) as fixed effects and a covariate of baseline value. The LS means and 95% confidence intervals for treatment difference and p-values for comparisons will be reported.

5.12.3. Device Questionnaire

Summary and analysis of device questionnaire will be made by treatment group using FAS.

For each item of device questionnaire in the first section, frequency and percentage of sum of "Strongly Agree" and "Agree" for Ateos, sum of "Strongly Agree" and "Agree" for Insulin Device, and "Both Devices are the same" will be presented by device at each scheduled visit.

Also, the overall preference for Ateos versus Insulin Device, which was evaluated by a percentage of sum of "Strongly Agree" and "Agree" for Ateos will be presented.They will be tested by the use of a one-sample binomial test using an exact method at each scheduled visit. The null hypothesis was that the preference for Ateos was equal to 50%, against the hypothesis that it was different from 50%.

For the summary of device questionnaire in the second section, the frequency and percentage of each category (i.e. "Rarely", "Sometimes", and "Always") will be presented by device at each scheduled visit.

5.13. Other Exploratory Variables

5.13.1. Dietary Evaluation

Summary statistics for total fats, total carbohydrates, total protein, and caloric intake amount (actual) by visit and change from baseline values will be provided using FAS.

5.14. Safety Analyses

Unless otherwise specified, the safety analysis set will be used for analyses of safety measurements. Also, p-values provided from all safety analyses will be used for guidance purposes only.

5.14.1. Adverse Events

An adverse event 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 classified according to the Medical Dictionary for Regulatory Activities (MedDRA) and summarized by treatment group. Adverse events will be summarized as treatment-emergent adverse events (TEAEs) defined as events that are newly reported or reported to worsen in severity after the initiation of study drug through the treatment period. The incidence of patients with at least one TEAE and the incidence of TEAEs by MedDRA preferred term and system organ class will be presented by treatment group. The incidence of patients with at least one TEAE that is considered to be possibly related to the study drug in the opinion of the investigator will also be summarized. Fisher's exact test will be used for treatment comparison.

SAEs and AEs leading to discontinuation of the study drug will also be summarized and listed.

5.14.2. Adverse Events of Special Interest (AESI)

The following adverse events of special interest (AESIs) will be summarized and listed. For AEs of interest associated with the thyroid gland (benign and malignant neoplasms), listing will only be provided.

- Allergic/hypersensitivity reactions
- Cardiovascular events
- Pancreatitis
- AEs of interest associated with the thyroid gland (benign and malignant neoplasms)

Because gastrointestinal AEs, such as nausea and vomiting, are among the most common events reported in patients treated with dulaglutide, summaries and analyses for time to onset, duration, and severity of nausea and vomiting will be provided.

5.14.3. Hypoglycemic Episodes

A listing of the individual hypoglycemic episodes by patient will be presented.

Summary reports will include both incidence and rates of hypoglycemia using safety analysis set. Hypoglycemia will be analyzed as "documented symptomatic," "asymptomatic," "probable," "severe," or "nocturnal" and for all events combined, "total hypoglycemia." See protocol section 7.4.1.2. Other categories, including the categories above defined with differing PG thresholds, may also be included in these analyses when deemed appropriate. The incidence of hypoglycemic episodes will be compared using Fisher's exact test.

Hypoglycemia rates will be summarized for periods of 1 year, 30 days, and 100 years (severe hypoglycemia only). The rate of severe hypoglycemia per 100 years will be compared between treatment groups using the empirical method (see Appendix 1). For each of the other categories of hypoglycemia, the number of hypoglycemia events during a specific period (rate) after randomization (for example, 0-16 weeks of treatment period) will be analyzed by using a negative binomial regression model. The model will include treatment and the baseline hypoglycemia rate (measured during lead-in) as a covariate. An offset defined as the log transformation of treatment exposure in the specific period (days)/365.25 days (or 30 days) will be included in the model to estimate the rate of hypoglycemia per year (or per 30 days). The proportion of patients with at least 1 hypoglycemic event in each category (incidence) during a specific period after randomization will be analyzed using a logistic regression model including treatment and baseline hypoglycemia rate value in the model.

Time to the first onset of total hypoglycemia and time to glucose-lowering intervention due to severe, persistent hyperglycemia will be analyzed between the groups using the semi-parametric proportional hazard regression model with treatment, baseline HbA1c (<8.5%, ≥8.5%), insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin) as fixed effects and baseline hypoglycemia rate (measured during lead-in) as a covariate. A Kaplan-Meier curve will be plotted for all treatments on the same graph, for presentation purposes. The proportion of patients that required intervention will be analyzed using Fisher's exact test.

5.14.4. Laboratory Analyses

All laboratory measurements including scheduled and unscheduled will be listed by patient by visit using all randomized patients. An additional listing will be presented for selected laboratory measurements that are outside the normal range. The normal ranges are defined by a lower limit of normal (LLN) and an upper limit of normal (ULN). A result that is greater than or equal to the LLN and less than or equal to the ULN is considered to be within the normal ranges.

Laboratory measurements will be summarized with respect to observed values and change from baseline by treatment group at each scheduled visit. For each analyte, the change from baseline to endpoint will be tested using Wilcoxon rank-sum test. Also, laboratory analytes with categorical responses will be summarized using frequency and percentage by visit and treatment group. Treatment comparisons will be performed using a Fisher's exact test.

The proportion of patients with treatment-emergent abnormal values defined by the outside of the normal ranges will also be summarized by treatment group and compared between treatment groups. The proportion will be calculated by dividing number of patients with abnormal value after baseline by number of patients with baseline and post-baseline values (patients who had abnormal baseline value in a same direction of interest will be excluded).

5.14.4.1. Amylase and Lipase

For amylase and lipase, the following variables will be analysed:

1) Number and percent of patients having measurements above ULN/2xULN/3xULN at screening (Visit 1), baseline (Visit 3), and each post-randomization visit (Non-cumulative and cumulative incidence)
2) Number and percentage of patients having measurements above ULN at baseline (Visit 3) with post-randomization values decreased from baseline at each visit.
3) Number and percentage of patients having measurements above ULN at baseline (Visit 3) with post-randomization values increased from baseline at each visit.
4) Number and percentage of patients having normal measures at baseline (Visit 3) with post-randomization values increased from baseline above ULN/2xULN/3xULN at each visit.

The cumulative number of patients having amylase above ULN at post-randomization visit is the number of patients having amylase above ULN at least once at any post-randomization visit (e.g. patients with repeatedly increased value are counted only once). The same holds for lipase.

5.14.4.2. Calcitonin

Threshold analyses will be performed for calcitonin. Some of these analyses involve a composite criterion involving change from baseline of at least 50%. For alignment with the calcitonin monitoring algorithm used in study, the baseline value for this analysis will be the mean of all nonmissing baseline observations. In cases where baseline is missing but postbaseline is available, the percent-change criterion will be assumed to be satisfied. Following threshold analyses will be provided:

- postbaseline calcitonin ≥20 pg/mL
- postbaseline calcitonin ≥35 pg/mL

Additionally, for each of these, a threshold analysis will be presented of patients also satisfying a change from baseline of ≥50%.

A shift table will be provided of postbaseline calcitonin to and from the following categories: (i) <20 pg/mL, (ii) ≥20 pg/mL and <35 pg/mL, (iii) ≥35 pg/mL.

5.14.4.3. Lipid

Actual value and percent change from baseline at each visit for lipid measurement will be summarized by treatment group using safety analysis set. For the percent change from baseline to Visit 7, wilcoxon rank-sum test will be conducted.

The lipid measurements to summarize in this study are “Cholesterol (milligram/deciliter)”, “Direct LDL Cholesterol (milligram/deciliter)”, “HDL Cholesterol (milligram/deciliter)”, “Triglycerides (milligram/deciliter)”. LDL cholesterol is also calculated using the Friedwald equation. In addition, LDL cholesterol is calculated using the Friedwald equation only for the patients with TG (Triglyceride) < 400mg/dL. For the patients does not have a direct LDL

cholesterol, but LDL cholesterol can be calculated with Friedwald equation, then the LDL cholesterol value based on the Friedwald equation will be included for this summary.

5.14.5. Physical Characteristics and Vital Signs

The vital signs include measurements of blood pressures, body weight and BMI.

Each patient will have their pulse rate and blood pressure measured 3 times. The measurements will be averaged separately for each patient at each visit; the average values will be used in the descriptive summaries and analyses.

Descriptive statistics for the actual measurements and change from baseline by visit and treatment group will be presented. For the change from baseline to Visit 7, wilcoxon rank-sum test will be conducted.

The incidence of patients with at least 1 measurement above the following levels after baseline will be summarized by frequency and percentage for each treatment group (Table GBGF 5.3). At the Visit 7, treatment comparisons between treatment groups will be performed using Fisher's exact test.

Table GBGF 5.3. Vitals Threshold for Study H9X-JE-GBGF

Systolic Blood Pressure (mm Hg)
≥140
Diastolic Blood Pressure (mm Hg)
≥90
Pulse Rate (beats per minute)
>100

5.14.6. Adjudication of safety events

Deaths, cardiac biomarkers, possible cases of pancreatitis, and non-fatal cardiovascular AEs will be adjudicated by an outside judgment committee external to Lilly. A listing of events that were sent for adjudication will be presented including the event as it was reported and how it was adjudicated.

5.14.7. Renal safety

The following biomarker will be assessed:

- Serum creatinine
- Estimated glomerular filtration rate (eGFR) estimated by Modification of Diet in Renal Disease (MDRD) formula using the coefficient for Japanese (Matsuo et al., 2009)
- Urinary albumin to creatinine ratio (UACR)

eGFR (mL/min/1.73 m^2) = 175*[serum creatinine$^{-1.154}$]*[age$^{-0.203}$]*[0.742 if female]*[0.808 if Japanese]

Mean and mean change from baseline to endpoint of serum creatinine, eGFR, and UACR by treatment will be presented. Proportions of patients shifting between CKD stages by eGFR from baseline by treatment will be presented.

Shift tables based on eGFR and UACR (Table GBGF 5.4 and Table GBGF 5.5) will be provided (CKD risk stratification in JSN 2012).

Table GBGF 5.4. CKD Stage based on eGFR for Study H9X-JE-GBGF

GFR Stage	eGFR (mL/min/1.73 m^2)
G1	≥90
G2	60≥ and <90
G3a	45≥ and <60
G3b	30≥ and <45
G4	15≥ and <30
G5	<15

Table GBGF 5.5. CKD Stage based on UACR for Study H9X-JE-GBGF

Albuminuria Stage	UACR (mg/g)
A1	<30
A2	30≥ and <300
A3	≥300

The following threshold analyses (with eGFR by MDRD formula and UACR) will be presented:

- Postbaseline eGFR <60 mL/min/1.73 m^2
- Postbaseline eGFR <60 mL/min/1.73 m^2, or UACR >300 mg/g

5.15. Subgroup Analyses

Subgroup analyses will be conducted for the variables of HbA1c, body weight, and hypoglycemia rate.

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

- Baseline HbA1c (<8.5%, ≥8.5%)
- Insulin regimen (basal insulin, premixed insulin, or basal/mealtime insulin).

For the subgroup analysis by insulin regimen, an analysis will be performed using an MMRM model or negative binomial regression model that includes the same effects given for primary or

secondary efficacy measures plus subgroup, subgroup-by-treatment interaction, subgroup-by-visit interaction, and subgroup-bytreatment-by-visit interaction. For the subgroup analysis by baseline HbA1c strata (<8.5%, ≥8.5%), baseline HbA1c will not be included as a covariate. The treatment-by-subgroup interaction at Week 16 will be evaluated using a significance level of 0.10, unadjusted.

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 CRP or CRS, or designee, will be consulted to determine whether it is necessary to amend the protocol.

6.2. Site Level Unblinding

This study includes a 16-week double-blind (dulaglutide versus placebo) primary treatment period, and a 36-week open-label extension treatment period.

To preserve the blinding of the study, the treatment assignments in the double-blind period will be blinded to patients and investigators until study completion.

Emergency unblinding for AEs may be performed through the IWRS, which may supplement or take the place of emergency codes generated by a computer drug-labeling system. This option may be used ONLY if the patient's well-being requires knowledge of the patient's treatment assignment. All actions resulting in an unblinding event are recorded and reported by the IWRS. If an investigator, site personnel performing assessments, or patient is unblinded, 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 or CRS for the patient to continue in the study.

In case of an emergency, the investigator has the sole responsibility for determining if unblinding of a patient's treatment assignment is warranted. Patient safety must always be the first consideration in making such a determination. If the investigator decides that unblinding is warranted, the investigator should make every effort to contact the Lilly CRP or CRS prior to unblinding a patient's treatment assignment unless this could delay emergency treatment of the patient. If a patient's treatment assignment is unblinded, Lilly must be notified immediately.

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. During this study, some members are unblinded (Table GBGF. 6.1). They do not have direct interaction with sites. They will not be allowed to join meetings which may affect the other members' blind condition, such as trail level safety review (TLSR) and data review meetings.

Table GBGF. 6.1 Unblinded Members through GBGF Study

Role	Reason	Data source for unblinding
Product Delivery personnel	To provide and manage clinical trial materials	e-CTS

Unblind Case Manager	To report SAE with unblinded information to Japan authority	e-CTS
Clinical Laboratory Operations	To manage and track sample shipping	CLRM

Abbreviations: CLRM = Clinical Laboratory Results Modernization; e-CTS = Enhanced Clinical Trials System; SAE = Serious Adverse Event.

7. References

Matsuo S, Imai E, Horio M et al. Revised equations for estimating GFR from serum creatinine in Japan. *Am. J. Kidney Dis*. 2009; 53:982-93.

[JSN] Japanese Society of Nephrology; Special issue: Clinical practice guidebook for diagnosis and treatment of chronic kidney disease 2012. *Nihon Jinzo Gakkai Shi*. 2012; 54(8):1034-191.

8. Appendices

Appendix 1. Empirical method for the rate of severe hypoglycemia

Traditionally, Poisson distribution has been assumed to draw inference for the rate of rare events. When the event is rare and the sample size is large, it is known that the overall number of events is approximately from Poisson distribution. However, for some not very rare events such as severe hypoglycemic events in T1DM patients, the total number of events may not be distributed from Poisson and may be over-dispersed. Assuming Poisson distribution may significantly underestimate the variance, and therefore may reduce the overage probability and inflate the Type-I error. Below we describe an empirical method in estimating the variance of event rate without assuming any distribution on the number of events.

Event rate without strata

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Event rate with strata or based on multiple studies

We use additional subscript "*k*" denote the corresponding variable for stratum (or study) *k*. The Mantel-Haenszel weighted estimate of λ across strata is

$$\hat{\lambda} = \frac{\sum_k w_k \hat{r}_{1k}}{\sum_k w_k \hat{r}_{0k}}$$

where $w_k = \left(T_{0,k}^{-1} + T_{1,k}^{-1}\right)^{-1}$. The variance of $\log(\hat{\lambda})$ is estimated as

$$\widehat{Var}(\log(\hat{\lambda})) = \frac{\sum_k w_k^2 \widehat{Var}(\hat{r}_{1k})}{(\sum_k w_k \hat{r}_{1k})^2} + \frac{\sum_k w_k^2 \widehat{Var}(\hat{r}_{0k})}{(\sum_k w_k \hat{r}_{0k})^2}$$

The 95% confidence interval for λ and the p-value can be constructed similarly using (1) and (2) as for the ratio rate without strata.

The event rate for each treatment group is

$$\hat{r}_i = \frac{\sum_k w_k \hat{r}_{ik}}{\sum_k w_k}, \qquad i = 0,1$$

and the variance estimation for the event rate is

$$\widehat{Var}(\hat{r}_i) = \frac{\sum_k w_k^2 \widehat{Var}(\hat{r}_{ik})}{(\sum_k w_k)^2}, \qquad i = 0,1$$

Event rate in the cross-over study (no period effect)

For 2 × 2 cross-over study, each patient may receive both treatments, e.g. control treatment in period 1 and experimental treatment in period 2. The above equation of event rate without strata for $\hat{r}_i$ can still be applied to the calculation of event rate for each treatment group but the variance of relative rate is different from $\widehat{Var}(\hat{\lambda})$ due to the variance within patient. To distinguished from $\hat{\lambda}$, $\tilde{\lambda}$ is used to indicate the relative rate of the experimental treatment versus the control treatment for the cross-over study

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

The variances of $\log(\tilde{\lambda})$ is

$$\begin{aligned}\widehat{Var}(\log(\tilde{\lambda})) &= \widehat{Var}(\log(\hat{r}_0)) + \widehat{Var}(\log(\hat{r}_1)) - 2\widehat{\mathrm{Cov}}\big(\log(\hat{r}_0),\log((\hat{r}_1)\big) \\ &= Y_0^{-2} n_0 S_0^2 + Y_1^{-2} n_1 S_1^2 - Y_0^{-1} Y_1^{-1} \sum_j \widehat{\mathrm{Cov}}\big(X_{0j}, X_{1j}\big)\end{aligned}$$

Let S_c^2 be the covariance of X_{0j} and X_{1j} for patient j and identical for all patient, and n^* be the number of patients who completed both the two cross-over periods, the variance of $\log(\tilde{\lambda})$ is

$$\widehat{Var}(\log(\tilde{\lambda})) = Y_0^{-2} n_0 S_0^2 + Y_1^{-2} n_1 S_1^2 - Y_0^{-1} Y_1^{-1} n^* S_c^2 \tag{3}$$

With the variance of in (3) $\log(\tilde{\lambda})$, the 95% confidence interval for $\tilde{\lambda}$ and the p-value can be constructed similarly using (1) and (2).